

RICE UNIVERSITY

Building Reasons Without Authority

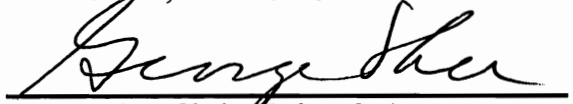
by

Stan Husi

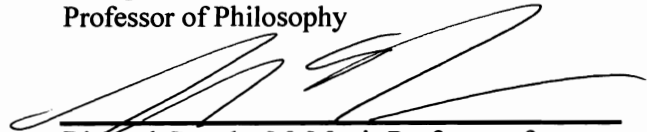
A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE



George Sher, Chair, Herbert S. Autrey
Professor of Philosophy



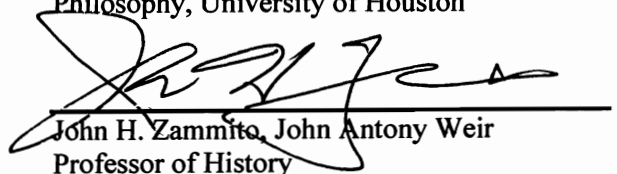
Richard Grandy, McManis Professor of
Philosophy



Alastair Norcross, Associate Professor of
Philosophy, University of Colorado at Boulder



David Phillips, Associate Professor of
Philosophy, University of Houston



John H. Zammito, John Antony Weir
Professor of History

HOUSTON, TEXAS

April 2011

Abstract

Building Reasons Without Authority

by

Stan Husi

My dissertation defends a comprehensive version of meta-normative skepticism which holds that no standard, norm, or principle has objective authority or normative force. The view does not deny either that there are norms, standards of correctness, and principles of various kinds or that it is possible both to succeed or fail in measuring up to their prerogatives. What it does deny is that any norm has the status of commanding with objective authority, the status of giving rise to objective normative reasons to take seriously and follow its demands. Many believe objective authority is required if we are to make sense of and explain the significance of our normative practices. Without authority, they fear, any critical standpoint vis-à-vis our practices would evaporate, even when we have reached a consensus regarding critical matters, which, without correctness, appears to reflect nothing but an ultimately arbitrary choice. I disagree, and argue that while authority cannot be accommodated within the world as we know it, we don't need it either. A chief goal of my dissertation is to propose a positive interpretation of our normative practices that dispenses with authoritative facts directing us what to do. The practical question of what to make of our practices and our involvement with them, I counter, retains significance only when pursued from an engaged rather than a

detached perspective – one that we adopt when, driven by our concerns and commitments, we actively participate in the resolution of practical problems, including the selection and development of norms to live by, searching for common ground for how to coordinate our individual and joint endeavors. Even though there are no definitive answers, this deliberative enterprise is not unconstrained; it is carried out within a tight web of norms that we do already accept, a web we continuously spin and expand.

Acknowledgments

I wish to thank my thesis committee: Alastair Norcross, Richard Grandy, David Phillips, and John Zammito. I wish to especially thank my advisor, George Sher, who has been absolutely wonderful and supportive throughout the years. I wish to thank my family, Linnet and Stanley, without whom I could not have done it.

I wish to thank Rice University, The Humanities Research Center at Rice University, and The Charlotte W. Newcombe Foundation for financial support.

Contents

Acknowledgments	iv
Contents	v
Preface	vii
Making Sense of the Meta-Normative Debate	1
1.1. Setting the Stage	1
1.2. Normativity and Norms	6
1.3. Norms and Evaluation	11
1.4. Norms and Favoring	13
1.5. Norms and Normativity	17
1.6. Norms, Morality and Rationality	20
1.7. Normativity and Authority	21
1.8. True Norms	23
1.9. Normativity and Reasons	25
1.10. Naturalism and Normativity	29
Against Non-Reductive Normative Realism.....	33
2.1. Introduction.....	33
2.2. Normative Non-Naturalism	36
2.3. What to Make of Realism – Impotent Metaphysical Arguments	50
2.4. The Epistemological Argument	58
2.5. The Practical Argument	71
2.6. The Transcendental Argument.....	86
2.7. The Refutation of the Transcendental Argument	90
Against Reductive Normative Realism	103
3.1. Introduction.....	103
3.2. Normative Naturalism	108
3.3. A Preliminary Case.....	116
3.4. Desire, Pure and Simple	118
3.5. Desire Cum Principle P	147

3.6. Value.....	162
3.7. Desire, Ideal and Informed.....	166
Against Agency-Based Accounts	182
4.1. Introduction.....	182
4.2. Methodological Preliminaries – What Constitutes What?	190
4.3. Agency, Shmagency.....	194
4.4. The Aim of Belief	198
4.5. Identifying the Constitutive Base for Action and Agency – An Initial Dilemma...	220
4.6. Constitutive Norms for Action: David Velleman’s Narrative	226
4.7. Constitutive Norms for Agency: The Integrity of Christine Korsgaard.....	240
4.8. The Past and the Future, Michael Bratman’s Plans and the Importance of George Sher’s Past	253
4.9. The Retreat.....	260
4.10. Inescapability and Authority	264
Practical Deliberation Without Authority	275
5.1. Introduction.....	275
5.2. Revisionism, Not Fictionalism	286
5.3. The Challenge from Deliberative Indispensability	301
5.4. The Outline of a Solution.....	306
5.5. So It Is All About Motivation?.....	312
5.6. Concern and Appealers	328
5.7. The Resilience of Concerns.....	340
5.8. Appraisal – The Problem	362
5.9. Appraisal – The Outline of a Solution (<i>or something close enough</i>).....	373
5.10. The Concern for Recognition.....	379
References.....	390

Preface

In this preface I would like to provide an overview of the dissertation. The first chapter *Two Concepts of Normativity* introduces the problematic of normative authority by insisting on the important distinction between norms and their status, between the formal feature of being directive in character and the substantive feature of directing with authority. The first distinguishes norms as norms, including their implicit standards of correctness, but it is the second that prompts the entire meta-normative problematic in the first place. There is a common tendency to under-appreciate the meta-normative problematic, and this is partly explained by our lack of an established vocabulary in which to express that problematic. The term *normative* itself is ambiguously used, at times denoting the directive element in norms, setting them apart from historical treatises and medical records, at other times denoting their authoritative standing, setting them apart from illegitimate norms. Presumably we are all realists about norms. The public arena in which we debate how to manage our individual and joint affairs is evidently characterized by a great plurality and diversity of norms and standards directing us what to do. Undoubtedly there are plenty of oughts and shoulds *according-to-norm-such-and-such*, plenty of opportunities to commit mistakes *according-to-norm-such-and-such*, and so forth. Language and the law exemplify that indisputable fact best. Whenever we open our mouths to form a sentence we engage in a norm-guided activity. And grammaticality is but one instance of for our thoroughgoing involvement with norms. Another is the law. It is unlawful in the United States to

hoist any flag higher than the Stars and Stripes. Yet whether authoritative reasons, not just officials, decree that I must comply in my own enclosed yard is unclear and represents a matter wholly distinct from the recognition that I would act contrary to the law if I do not. The normative problematic as I envision it cannot be appreciated solely in terms of standards of correctness, but only in terms of the authority of standard of correctness. Norms are not the solution, but the problem.

The second chapter, “Against Non-reductive Realism,” challenges the idea that some norms just are authoritative practice-independently, even though there is no explanation why. This form of non-reductive normative realism has recently gained considerable support, and holds, in David Enoch’s succinct summary, that *there are response-independent, non-natural, irreducibly normative truths, perfectly objective and universal ones, that when successful in our normative inquiries we discover rather than create or construct*. Yet to postulate an independent normative realm, I argue, presupposes a metaphysics that systematically undermines all epistemic access to normative truths and renders it mysterious why once detached from our concerns they should practically matter to us in the first place. If certain norms were authorized in practice-independent Platonic heaven, how should we ever find out which, and why should we obsequiously follow *them* rather than those that reflect our *own* concerns? Moreover, the strategy most prominently cited in support of non-reductive normative realism, namely transcendental argumentation, is unsuccessful, even if rhetorically quite impressive. The argument begins by pointing out that settling what to accept, and in particular whether to accept a skeptical position such as my own, appears to be a norm-driven enterprise. Yet, it

continues, skeptics cannot coherently view as authoritative the norms *they* need to rely upon in advancing their case without also relinquishing their very skeptical denial. How, then, can they coherently recommend their view? In response, I admit that making the case for and against any hypothesis does essentially involve norms: we need criteria of argumentative correctness and success. Yet what we do not need is to presume that the norms underwriting philosophical argumentation are equipped with practice-external authority. We only need norms that in fact facilitate our epistemic and dialectical ends, and their employment is not bound up with or contingent upon the metaphysics of Platonic realism.

The third chapter, “Against Reductive Realism,” rejects the proposal that normative authority can be explained on the basis of our desires and endorsements. Initially, desire-based accounts have a lot going for them. The issue of a norm’s authority arises in connection with agents, the unique consumers of standards and reasons. They are the ones to whom norms must be addressed, and it is they who must determine whether to take seriously what is so addressed. This naturally leads to the thought that a norm’s authority consists in its voluntary endorsement and endorsement, the proposal must say, is a complex form of desire, a disposition to do something, instead of a judgment on the norm’s authority (if it were such a judgment, the proposal would move in a circle). Yet desire-based accounts face a fatal dilemma. If the relationship between agents, their desires, and their choices is understood purely descriptively, as a complex natural-psychological phenomenon, it is clear nothing normative can fall out of it. If, however, we appeal to additional normative principles, which asserts that agents ought or have reasons to further

their desired ends to explain the normative significance of desire, then we have essentially introduced a principle whose authority cannot itself be accounted for in terms of desire. And this, I argue, involves costs for desire-based accounts that far exceed the mere admission of a gap in explanation. It amounts to no less than the rejection of the very rationale that motivates desire-based accounts in the first place.

The fourth chapter, "Against Agency-Based Accounts," introduces and rejects agency-based accounts of authority. The basic idea is this: For a norm to be authoritative is for it to be constitutive of our own agency and capacity to make practical choices. A norm that enables you to be who you are and that empowers you to ponder what to do must represent a standard you cannot escape. This quintessential Kantian strategy concurs with my own approach in its emphasis on the vantage point of the deliberating agent. The problem, however, is that constitution and authority represent rather distinct phenomena, and to equate them is to equate apples and oranges. At the very least, constitution cannot be the whole story on authority even if it is part of it. This becomes apparent once we realize that many norms that are held to be authoritative are not constitutive of anything. More importantly, though, the kind of necessity that underwrites constitution is not the same as that underwriting normative authority. Constitution concerns what we can't help doing more than what we should be doing. It belongs more to the purview of the engineer who designs complex norm-consuming systems than to that of the ethicist and practical philosopher. Suppose acceptance of norm N turns out to be constitutive for doing X. In that case you cannot do X while disrespecting N. Should

you also care about doing X, then you would only achieve what you care about so long as you implement N. *You gotta do what you gotta do*. Yet nothing more seems to follow. In particular, it does not follow that you normatively must comply with N or that you have any reasons to do so. The impossibility of *doing-X-while-disregarding-N* could reflect some basic constraints in design-space akin to that one cannot build stable bridges while disrespecting the laws of gravity. Moreover, the inescapability of doing X itself would not change the normative situation either, but solely place yet another constraint on your practical options. Thus, even if you absolutely had to do X – a choice you simply could not evade – and further that doing X was required for complying with N, this hardly would entail any reasons to comply with N on your part. One constraint would lead to another, but the fact remains that being constrained is not the same as having reasons. You merely would find yourself trapped in a tight corner. Might does not make right, as normative force evidently differs from brute force.

The fifth chapter, “Reasoning without Authority,” develops my counterproposal explaining how norms and social practices can have significance for us without appealing to practice-independent authority> I call the resulting view “revisionary subjectivism” and present deliberation as a very sophisticated form of norm-guided motivation with the aim of accommodating core features of practical deliberation. The chapter starts out with a discussion of the fictionalist response to skepticism, the view that we may manage to retain the benefits of knowingly discredited practices by downgrading our epistemic attitudes towards them from *believing* to *making-believe*. Against this I show the fictionalist response to be

inferior to the revisionist alternative, for once fully worked out, fictionalism is going to lead us right to the very doorsteps of revisionism. The benefit the fictional attitude is capable of affording is due to the retainable elements of the corresponding practice. If so, there always is an available revisionary alternative which “shrinks” the original practice to precisely those retainable elements, and which has the advantage of clear-headedly dispensing with the barely-stable fictionalist attitude. Because of this we should stop *making-believe* in what does not work and start *believing* in what does work.

The chapter then addresses the fundamental challenge that without a normative account there cannot be an adequate account of deliberation, the challenge, in particular, that revisionary subjectivism cannot accommodate the notion of correct deliberation. To this I respond by aiming to show how the subjectivist can accommodate correctness-permitting deliberation from within by helping himself to norms underwriting the deliberative process. These are norms void of normative authority, yet they nonetheless generate formal correctness conditions which, in conjunction with motivational force acquired through our commitments, enable a form of practical deliberation incorporating internal correctness and incorrectness conditions. After this, the chapter turns to a close discussion of concerns, the core component engendering deliberation according to revisionary subjectivism. Attending to concerns more closely reveals an astonishing degree of complexity and richness, which, despite the long noticed importance of desires and concerns for practical deliberation, has not always received adequate treatment in the literature. Concerns are unruly fellows, displaying an intriguing set

of properties including diversity, specificity, particularity, and synchronic and diachronic stability. Moreover, concerns appear to contain an element of appraisal, where there often seems, in the relevant object, an element of invitation *drawing in* the concern: concerns do not present themselves as a mere blind reaction aroused by certain aspects of their objects, but as a receptivity to their *attractiveness*. This sometimes enables us to construe a concern as *called-for by the attractiveness of the relevant feature*, as something that can be *appropriate or inappropriate* depending on whether the feature is in fact attractive. In light of this, one of my objectives in the chapter is accommodate the element of appraisal in a manner consistent with skepticism.

This final and positive part of the project supports the first and negative one by neutralizing the general worry that our practices would naturally falter without an external authority supporting them. The case for practice-independent authority rests on painting a grim picture of the alternative. Yet if a brighter picture of that alternative can be painted, the case for such authority loses one major column of support. This is exactly what I set out to do in the final chapter. The idea of normative practices without authority is intriguing because it liberates practical thought from all doubts about an authority transcending our practices. It allows us to grow content with thinking of practical thought and its object as fundamentally of our own making – the only way that makes sense to me. What I share with most of my opponents is the core belief that our practices and their internal standards are indispensable. The lesson I draw from this is different, however. What I wonder is how such a central piece of our lives could be held hostage to uncertain

metaphysical truths. Rather than joining into the defense of our practices by defending authority, I maintain no such defense is needed to begin with. My opponents serve our practices poorly by rendering the very point of them contingent on normative metaphysics. By revising the interpretation I seek to conserve the practices.

Chapter 1

Making Sense of the Meta-Normative Debate

1.1. Setting the Stage

This is an essay in normative metaphysics. It investigates the nature, reality and practical significance of normativity. The thesis I defend combines a skeptical stance on the existence of normativity with an equally skeptical stance on its practical significance. Normativity, I argue, cannot be accommodated within the natural world as we know it; I consider my naturalistic commitments to exclude normativity from my ontological repertoire. It is the *combination* of ontological *and* practical skepticism about normativity, however, that promises to present the skeptical position in a novel and favorable light overall. My denial of the normative not only is attenuated by my corresponding denial of its practical significance, or at any rate by my suggestion that its practical significance may be renegotiated; it also receives support by neutralizing the

perhaps most powerful counterargument. Many have forcefully argued that “normative” discourse would naturally falter without normativity supporting it; their case for normative realism partially rests on painting rather grim a picture of the alternative. Yet if a brighter picture of that alternative can be painted, the case for normative realism loses one major column of support; the case I wish to make, then, is that far from being inevitable, the realists’ portrayal of irrealism actually constitutes a rather dubious sort of fear-mongering.

Moreover, we hardly know what that alternative would look like to begin with; there are several systematic treatises defending normative realism, but very few, if any, opposing it; the sketchy portrayals of normative irrealism offered by realists are all we really have – teeming with confusions and misunderstandings. We certainly find plenty refutations of particular normative domains – most prominently the moral one – but as we shall see that’s a different story. So without having a systematic investigation of normative irrealism at our disposal, the basis for favoring realism over irrealism must appear somewhat imbalanced. Even those who disagree with irrealism may welcome to learn more about what exactly they are disagreeing with. With this appeal to their curiosity, I’d like to invite them to imagine normative discourse without normativity, and possibly to be surprised. If my case is sound, not only is normativity the kind of suspicious property many have suspected, but also rather dispensable. Normativity, put bluntly, is something we neither have nor need.

Normative discourse without normativity – such an idea promises an intriguing vista on practical thought and discourse, I believe; not only would it immunize practical

thought against an increasing number of nagging doubts about the normative; it would also allow us to grow content with thinking of practical thought and its object as fundamentally of our own making – the only way that makes sense to me; content because practical thought would emerge not any longer as wanting or deficient solely because it was of our own making. Putting my cards on the table, then, this project is in part motivated by the intention to bolster a broadly constructivist agenda in political philosophy by neutralizing one of its greatest worries: that constructivism is deficient for its inability to account for normativity; I agree that it is unable to do that, but I disagree that it needs to do that. If accounting for normativity presents an unrealistic demand, we may rather want to reconsider that demand; *unrealistic*, because it would be a funny demand indeed to account for that which does not exist. And we already know a few of such unrealistic demands that have at times been made of practical thought or parts thereof – that there must be a God, mostly, without whom many feared there would be no right and wrong; and yet, some of us got over that. In fact, I believe asking for normativity is not completely unlike asking for a God; both seem like asking for the ultimate backup, for that which with unquestionable authority would license the way we conduct our lives alone and in company; but we need no backup and no license; we can learn to live without that which we cannot have: normativity.

But how can the union of normative discourse and normativity be even so much as optional? Normativity appears to be part and parcel of normative thought; moreover, the request for normativity is profoundly motivated by elements inherent in normative thought, as I will show shortly. Yet however firm their current union may be, it can still come apart. To show how, I want to introduce an important distinction, namely between

practices and *interpretation of practices*; we engage in normative thought, and we interpret that engagement; we think about what to do, and we have a conception of what we are then thinking about. The normative interpretation – the one I wish to dispense with – has it that in thinking what to do we are responding to what we are called upon to do; practical questions supposedly receive answers in authoritative facts of what is to be done – facts, I claim, that do not exist. Once interpretation is distinguished from that which is interpreted, it becomes possible to replace one interpretation for another; a chief goal of this essay, then, is to devise such an alternative interpretation of practical thought that is not in need of normative facts. What I share with most normative realists is the core belief that practical thought is indispensable; the lesson I draw from this is different, however; what I wonder is how such a central piece of our lives could be held hostage to uncertain metaphysical truths. Rather than joining into the defense of practical discourse by defending normativity, I plea no such defense is needed to begin with. Realists serve our practices poorly by rendering the very point of them contingent on normative metaphysics. By revising the interpretation I seek to conserve the practice.

Similar interpretative shifts have occurred before; picking up on my analogy from religion again, morality was long perceived to depend on divine authority, but is not anymore. Still, how profound the conviction at times was is witnessed by the allegiance it found even in thinkers who allegedly challenged it. Grotius – who contended that morality could conceivably have a certain degree of validity without God (Schneewind 1998:67-68) – thought that strictly speaking “duty and obligation ... necessarily supposes a superior power, a supreme master of mankind.” (Schneewind 1998:73) Kant seemed to concur. In his critique of rational theology, Kant argues that God’s existence is to be

postulated as a condition of what ought to be. Having God's existence in mind, Kant writes: "Since there are practical laws that are absolutely necessary (the moral laws), then if these necessarily presuppose any existence as the condition of the possibility of their binding force, this existence has to be postulated." (Guyer 2006:152, A 633-4) Dostoevsky, of course, put the doctrine best when he ventured that God's death would result in unlimited moral permissibility. Moral thought, then, was long thought to target at God's demands; no such demands, no target and no point, or so many must have thought. Yet our conception of moral thought certainly has changed and evolved, without even altering much of moral thought itself; we still believe murder is wrong, even if we aren't told so by God. The normative interpretation of normative thought, I thus propose, may turn out just as optional as the theological interpretation of morality.

But what is normativity anyway? To answer this question, I will propose a norm-based approach; to a first approximation, I take normativity to be a property that distinguishes norms with authority from those without. I believe this approach is particularly fruitful, for three reasons; first, it allows us to tackle the notion of normativity step by step; thereby we can expect to gain greater illumination than if we took on the notion as a single piece; secondly, I think the approach is intuitively appealing. There are a great number of norms that have been or could have been issued on us; yet at best we only take a small number of them to be authoritative; there is an intuitive distinction between norms that merely issue claims on us versus those that do have claims on us.

Thirdly, and most importantly, the approach actually offers a motivation and not just an analysis of the idea of normativity. It opens up a perspective on normativity that

would explain why the idea may have been introduced in the first place; it presents the idea of normativity as a response to a problem and not merely as a product of our creative minds. The norm-based approach, then, does not emerge as a mere exercise in conceptual analysis, which is a most welcome result; unmotivated conceptual analysis of *normativity* makes for a questionable enterprise, I believe, as it immediately grapples us with thorny questions such as whose concept we are analyzing anyway and what significance the concepts we'd end up with could possibly have (normativity – schormativity). Motivation as opposed to analysis would render it simply unnecessary to settle which of the many definitions of normativity got it “right.” And so while I do believe that I can show my conceptual choice to correspond to important parts of the philosophical literature, what I'm ultimately interested in is the philosophical problem to which I take the idea of normativity to be a response, and not the mere idea itself. That's why I believe the norm-based approach is so attractive, for it nicely explains why anyone could have been interested in normativity in the first place.

1.2. Normativity and Norms

I take it to be a minimal yet distinctive feature of normative domains or categories that they open up space for success and failure; it must be possible for something to go wrong. A normative category must not only give rise to some distinction between what does and does not satisfy or belong to that category; any old category does that. Rather, it must formally encapsulate an unfavorable stance towards instances that do not satisfy that category. Such instances must count as nonconforming as opposed to merely falling

outside the category's application or extension. In fact, it is only because a normative category does apply to nonconforming instances that those instances can be classified as nonconforming in the first place. So whereas the relationship between non-normative categories and their instances is exhaustively captured in terms of formal satisfaction, the relationship between normative categories and their instances is importantly different and more complex.

The difference is most apparent when we consider negative cases of dissatisfaction or nonconformity; take ELEPHANT, a non-normative category, and HEALTHY, a potentially normative one. Things that aren't elephants are in no way wanting; they just aren't elephants. It would be an odd way of speaking to say that a mouse fails to live up to the category ELEPHANT. Unhealthy things, in contrast, can sensibly be thought of as failing to live up to being HEALTHY. HEALTHY, I said, is *potentially* normative, as it is not necessarily used that way in every context; one may always abstract from the normative aspect of a normative category and focus exclusively on its descriptive aspect; *HEALTHY* used this way would just be another distinction in how things could be. The normative aspect of categories by no means preempts them from being used merely descriptively; still, it is once we can sensibly think of negative instances as not merely dissatisfying, but as failing to conform or live up to some category that this category becomes normative.

This brings us to norms, then; norms by their very nature encapsulate conditions of success and failure, and thus provide an excellent entry for thinking about normativity. Norms, I suggest, are models for a range of objects. The full specification of a norm

would encapsulate at least two elements. First, each norm would specify a range of objects it applies to; norms always are norms *for* something, and cannot be empty or lack ‘subjects’ to be governed. In turn, only that which is within a norm’s range can sensibly be judged in terms of that norm; norms for elementary school students do not apply to high school students; not because high school students are perfect can they never violate elementary school norms; they cannot do so simply because they are outside the range of elementary school norms; I will call a norm’s range of application its *application* or *jurisdiction*. Notice that no principal limits have been placed on what could count as a norm’s application; norms often apply to people and their behavior, but might just as well apply to artifacts, hypothetical scenarios, or the world as a whole.

Secondly, norms specify models for how the objects within their applications shall be. Norms encapsulate a directive element. Figuratively speaking, norms are maps with a particular spin or direction of fit, as they do not map what is but what should be. Anscombe’s (1967) frequently used example of two “identical” shopping lists used for different purposes comes in handily; the list that directs the father what to buy is a norm; the list of what he did buy compiled by the detective afterwards is not. Or take another example; architectural blueprints are norms, while architectural digests are not; blueprints specify how a list of materials is to be arranged, digests how they were arranged. Models are complex units encoding information, and one cannot tell whether they are used as norms or representations solely by considering the information encoded; whether or not some model functions as a directive or representation also depends on how a misfit between model and world would have to be construed. If in such a misfitting case “the fault would be placed on the world,” the model would be used as a directive; otherwise as

a representation. Norms call upon their application to be as specified by the model, representations are called upon to specify things as they are. This characterization suggests that the distinction may not be as straightforward as it initially appeared; representations as such may also be norm-governed, possibly on a higher level: what makes a representation a representation is that it is to accord with how things are, which clearly has the form of a directive. It is not easy to make all this more precise and to replace the figurative with the non-figurative; still, it is quite familiar; it may be hard to come up with a better characterization of norms; but most would probably agree that norms as such, including their existence and epistemological status, are not terribly problematic. In any case, we better ensure that what may be said about norms in the abstract comports with the fact that norms are both prevalent and familiar.

Thus, instead of further embarking upon a detailed treatment of norms, I wish to plea to the familiar. We can readily identify an abundance of norms operating in virtually any corner of our lives. Norms are as common as what they govern – which is basically everything; even talking about norms is recognizably norm governed – if only because language is; hence no later than when we start learning our language – figuring out how to distinguish correct from incorrect usages of that language – are we immersed in norm-governed behavior. We certainly cannot always articulate the norms we are guided by; making norms explicit requires some effort. Still, that there are norms for most of our doing is as plain as it could be. Just think about the familiar academic cocktail party. There are norms concerning what to wear; when to arrive; how to eat, and what to eat; what to talk about and what to only hint at; what jokes to make, and how to react to others' jokes that one may not make; how long to stay, and when to leave; who to say

goodbye to and whom to thank; what to recall about the party later on and what to leave as that which is not to be talked about; whether and when to return the favor; and then, how and when to host the next party, for which there will be norms again concerning such matters as what to wear, and when to arrive...

I also believe my focus on norms is compatible with a wide array of philosophical traditions; more controversial material will emerge soon when I turn from norms as such to questions concerning their status; for now, I would like to set up the point of departure so as to welcome most everyone to join the trip. Norms are marvelously versatile and can come in all shapes and sizes; norms can be principled or unprincipled, general or particular, atomistic or holistic, categorical or hypothetical; norms can specify conditions tight or loose for their application, address affairs large and general or tiny and concrete; norms can be implicit or explicit, encoded or non-encoded – perhaps even unencodable. Furthermore, norms can be structured in all sorts of ways; most norms are not isolated, but rather come organized in form of entire webs of norms; norms are often nested in each other and are governed by higher order norms; there are plenty of norms governing how to devise or set up norms, how to debate norms, how to incorporate norms, even how to follow norms, etc. There are norms for when to make exceptions, norms for when to suspend norms, and so forth. In light of such flexibility, we should be able to find norms suiting every kind of philosophical predilection and background.

Let me show this for a prominent few. Principles are norms, and consequently, principle-based traditions such as consequentialism and deontology should have little troubles countenancing the importance of norms. What about traditions that reject the

primacy of principles? No problem there, I think; while principles are norms, norms need not be principles; virtue-oriented traditions employ notions such as what kind of person to be or whom to be angry at in the right manner at the right time and so on and so forth – and it's not hard to detect a norm in all of that.

1.3. Norms and Evaluation

Let me expand my norm-based approach to the domain of evaluation and value. I believe evaluation or evaluative statements will fall into place smoothly and easily once we identify their norm-based character. Evaluative statements, I suggest, are statements about particular instances judged in light of some background norm. If such an instance is within that norm's application, the evaluative statement will be true or false depending on whether that instance complies with the model specified by the norm; if it does, that instance would be valuable along some dimension of value – evaluations, after all, are statements concerning value. The full explication of the content of an evaluative statement would involve the specification of a norm; every norm would in turn give rise to a potential infinity of true evaluative statements, relative to that norm.

Consider the norm “one must not wear green socks to dinner parties;” let's use the term *socky* for behavior that violates this norm; an evaluative statement such as “Peter behaved sockily again” would be a true (negative) evaluation of Peter's behavior if he indeed was wearing green socks to a dinner party. The norm-based account of evaluation admittedly is somewhat deflationary; the truth of evaluative statements is easily secured; all that is required is a norm operative in the background with which we can hook up the

relevant evaluative statement, and a sample of plain descriptive facts. Evaluative facts piggy bag on facts about norms and descriptive facts; as a consequence, whatever can be said about norms will have implications for evaluative statements as well.

The norm-based account of evaluation or value has various attractive features. First, it explains why value is inherently dynamic; value does not sit still, but engages our active capacities; value is not merely another aspect of the world to be taken notice of – it engages more than our representational capacities; it bears a direct and dynamic relationship to what is to be done. Since norms are directive in nature, the norm-based approach of value can do justice to this dynamic element of value. Secondly, the approach squares well with a wide variety of prominent accounts of value; take, for instance, Scanlon's buck-passing account of value. On Scanlon's account, something is valuable roughly if it calls for a certain response; a piece of art is beautiful, say, if there are reasons to admire it. Now, as will become clear later on, I cannot go with Scanlon all the way, but here's how far I can go. Calling for a certain response is a norm-governed affair. The statement that a piece of art is beautiful can be understood (in part) as involving a norm; a norm calling for admiration with regard to that piece. Or take response-dependent accounts of value; the responses taken to be constitutive of value must all have a directive direction of fit; they thus could qualify as mini-norms of sorts; my desire to be a good philosopher is a norm calling on me to be a good philosopher, a norm I endorse. Or take perfectionist accounts; again, what does or does not qualify as the suitably perfectionist state can well be stated in terms of norms; if one is to exercise regularly in order to reach perfection, the corresponding norm calling for regular exercise could be taken to (partly) specify the content of that perfection.

Thirdly, the norm-based account of evaluation would nicely explain how and why evaluation and directives so intimately hang together. An example from the moral domain will make this plain: What we want to say are things such as that it is because some action would produce bad effects that one must not do it. Now, the question is what explains that, as most would think, the badness of those effects and the prohibition against doing it are not just accidentally, but inherently related? Few would believe that it was just an accident that one must do what has value rather than disvalue. The norm-based account of evaluation, I think, has a ready answer here. According to that account, evaluation and directives are really just two sides of the same coin. For something to be good, as the norm-based account has it, is for it to be as it is supposed to be; and how it is supposed to be is specified by some norm; the norm says *be so-and-so*, and it is so-and-so – and hence it is good. We consequently do not need an extra principle connecting *something is valuable* to *be like this!* Being valuable already couples *Be like this!* with *being like this*.

1.4. Norms and Favoring

The norm-based approach, I think, can also be employed for getting a firmer handle on the notion of *favoring*, an important one in contemporary meta-ethics; many believe, for instance, that favoring is what reasons characteristically do. What I want to suggest now is that favoring is a norm-supported relationship as well; the full specification of a particular favoring relationship would make reference to some norm or other. Let's take an example. The fact that walnuts are healthy due to their richness in

Omega 3 fatty acids favors eating them. Why? The simple answer is that one should eat healthy things (a norm) and walnuts are healthy; a more complex answer would perhaps specify a variety of norms interacting with each other in this context; if I'm correct, the evaluation *healthy* already contains a norm – a norm for how food is supposed to be; thus, the norm for what to eat and that which underlies healthy would already be related; eat that which is fitting for eating, where fitting for eating is that which supplies the body with what it needs so that it can function as it is supposed to function (another norm), and so forth.

The norm-based approach to favoring suggests the following picture, then; the favoring-relationship holds between some states of affairs and a response in virtue of a supporting norm. Whether or not we want to think of the relevant norm as participating in that relationship or merely sustaining or enabling it I do not care; a lot discussion has been spilled lately on the distinction between favorers and enablers – a distinction I'm not sure carries much substantial weight; for the record, I'm fully content with the weaker construal where norms need not be part of the favoring relationship itself; as long as norms play a crucial job in explaining that relationship, I'm fine, whatever the exact job description may turn out to be.

The important point is that it is always a fair question why some state of affairs favors a particular response; by answering that question we bring out what norm is operative for sustaining that favoring relationship; why does the fact that walnuts are healthy favors eating them? The simple answer, that this is so because one should eat healthy things, immediately brings us to a norm. That *favoring* is norm-supported is often

disguised. For one thing, as I argued above, many favoring-statements do already cover norms; this generates the appearance that no norm needs to be mentioned in order to explain some favoring, overlooking that a norm has been in play all along. But also, many norms that support some favoring or other simply are taken for granted, even trivial. Paying them any extra attention seems superfluous; still, trivial norms are norms nonetheless.

My thesis, then, is that states of affairs cannot be favorers just by themselves; favoring is like calling for something, and states of affairs don't call for anything; they are fully satisfied with how they are; there is a gap between the mere existence of some state of affairs and having favoring powers; that gap needs to be closed and norms do just that. Some disagree, however, and argue that nothing additional is needed to close that gap; the specification of certain states of affairs, they think, can be explanatorily complete as the favoring powers of those states of affairs are concerned; they hold the relationship between the existence and favoring powers of some state of affairs to be seamless.

Three reasons speak against that view. First, as I already explained, the appearance that in many cases favoring needs no norms is faulty for it disguises the fact that norms are already contained in the relevant evaluations. Many descriptions of states of affairs are anything but innocent as they contain plenty of norm-laden terms; if *beautiful* involves a norm calling for admiration, then mentioning Venice' beauty is indeed sufficient to settle whether it favors admiration. This and other value-laden

descriptions of states of affairs thus do not provide counterexamples to the claim that every favoring involves a norm.

Secondly, on a broader semantic level, the view strikes me as utterly implausible in itself. Applying Moore's open question argument to the case at hand, I believe it always remains an open issue whether or not some state of affairs favors even after one has come to full comprehension of that states of affairs. One could wonder about this matter without being factually confused or partially ignorant. The coldhearted dissector of reality who is at a loss of finding any favoring need not be charged with having done incomplete work; he merely failed to look at things through the classes of norms. This suggests – if only defeasibly – that the existence and the favoring of some state of affairs can be dissociated.

Thirdly, if favoring was fully inherent in states of affairs, it would be hard to explain certain sorts of variability; whether or not, what, and how strongly certain states of affairs favor often depends systematically on external factors; but how can matters concerning favoring both be sufficiently accounted for by internal factors of certain states of affairs and yet be explanatorily dependent on external factors as well? Consider again Venice's beauty. Under certain circumstances, this fact favors visiting Venice – a response that let's suppose is not already contained in describing Venice as beautiful. The emphasis, however, is on 'certain circumstances.' Suppose Venice' very existence is at stake; the city will fall if any more visitors enter it. In that case, what Venice' beauty favors is not visiting it. Yet the fact that Venice is beautiful has not been changed in the least; all that has been altered are external circumstances, producing the result that the

identical states of affairs now give rise to opposite favoring powers. Moreover, it is important to stress that in both cases of favoring the relevant factor remains the same – Venice’s beauty; if Venice was as ugly as adjacent Maestro, I’d say go for it – considering only aesthetic reasons, of course. Yet if favoring was inherent to states of affairs, wouldn’t that suggest that Venice’s beauty contains two favorings with contradictory aims – to visit it and not to visit it? However this may be, the norm-based approach offers a solution that would silence all these worries in one stroke; it explains why those and only those states of affairs that favor do favor; it also explains how such changes in polarity can occur due to the fact that different norms are operative in different circumstances. If Venice is safe, the relevant norm is to visit beauty. If in danger, the relevant norm is to protect beauty; without including those norms in the explanatory basis for why and what Venice favors, such sudden shifts in favoring powers would remain rather mysterious.

1.5. Norms and Normativity

Let’s turn now to the status of norms; I believe that’s where all the action is. If norms govern all corners of life, the question is with what right; their status must become a rather intriguing issue. Norms, I suggest, provide an excellent entry for thinking about normativity because they introduce *normativity* as a response to a problem created by norms. Norms are the starting point, but not the end point, for thinking about normativity.

Many seem to think otherwise; speaking of the moral ought, the quintessential normative province, James Griffin, writes: “To say that something ‘out to be’ is to say

that it is what conforms to some standard, norm, or regularity. ... 'Moral oughts' are no different. They claim merely that there is a norm in the background to which a certain action would conform." (Griffin 1996:81) Yet having a norm in the background is easily secured; so easily, in fact, that one can readily understand the discomfort this would create for those who feel that having a norm in place cannot be all there is to normativity. There is a great number of social systems sustaining an even greater number of moral and non-moral norms in the background; yet we do not think that just any of those norms is worth heeding; anthropologists have identified and studied a huge variety of norms in the cultural record, many of them abhorrent. They found norms prescribing murdering and sacrificing of outsiders, systematically discriminating women and minorities, permitting slavery and intolerance, etc. In many societies men are still called upon to dictate women's fates without residue. For sure, the questionable status of norms is felt strongest when it comes to norms far afield; but even when we turn to norms operative in our own society, most of us find some norms we see little point in complying with, if we don't outright reject them. My favorite example are norms I believe erected to perpetrate a state of massive inequality; norms, for instance, that effectively inhibit any redistribution of goods, even if without fault of their own some have everything and others nothing. Yet in such opposition we need not and usually do not deny the existence of what we are opposing; in fact, only if the existence of such norms is taken for granted does the opposition have a real target and some bite. Griffin's conception which takes as the only requirement for the truth of some ought statement that some norm exists in the background is surprisingly weak; it must fall short of capturing even in principle the

intuitive difference of norms we ought to comply with and those we do not, since plainly enough norms of both kinds do exist in the background.

Indeed, if the problem of normativity was to find ‘norms in the background,’ we could happily report “problem solved;” we may further inquire into semantic or sociological aspects of norms. But it can hardly be the existence of norms that philosophers have worried about for so long. If there is a problem stemming from such existential matters, it is because there is too much rather than too little of it. There are so many norms, many of them in mutual conflict; we cannot consequently take all of them seriously – but how to select those we should? And what would it be that recommends some norms over others? It is not for their shortage, then, but rather for their abundance that one starts to wonder what status particular norms or even norms in general may have. If one thinks that norms in the background is all there is to normativity, I’d like to ask with respect to some more controversial norm whether in skeptical moments the assurance of the existence of that norm is all one wishes to be assured about. If indeed the existence of that norm is what causes one to worry, one may just make it up and have it.

Moreover, the existence of norms cannot be thought of as a requisite for normativity either; the norms that are truly worth heeding might not have been installed or thought of yet; those norms may not yet fill the background. And if so, this altogether contingent matter by itself is unlikely to disqualify any of these ‘non-existing’ norms. I conclude, then, that the existence of norms is not what constitutes normativity; simply having norms in the background is not enough. The idea of normativity is the response to the fundamental problem of which of all those norms that fill or could fill the background

are in fact authoritative; normativity is that generic feature that distinguishes authoritative norms from mere pretenders.

1.6. Norms, Morality and Rationality

If the notion of normativity is a response to a generic problem created by norms, the notion will be as general as the problem. Let me stress that generality. What I seek to understand is not how this norm can be normative if that norm is; rather, the problem is what could make any norm normative. Many believe the normativity of certain norms is less problematic than that of others; there is a huge literature trying to explain the normativity of morality in terms of the normativity of rationality; or that of other-regarding demands on the basis of self-regarding demands. If there is any such asymmetry, it would be good advice to focus on the least problematic case and explain what makes that norm normative. Presumably, that case would be some norm of self-interested or instrumental rationality. I doubt, however, that there is any such asymmetry. Once one appreciates the general nature of the problem, the notion that the normativity of certain norms can be taken for granted while that of others requires explanation seems rather silly. If there is a problem with normativity, as most would grant, then it is a problem across the board. It is just as big a conundrum to explain what could make egoistic norms normative as what could do the same for altruistic norms; the normativity of moral norms per se seems no more troubling than those of purely self-centered rationality; norms of the first sort demand to heed the interests of others; those of the second sort demand to never heed the interests of others unless there's something in for

me; these are just two different norms; that they differ in content by itself does not explain why one is authoritative while the other isn't. Explaining the normativity of morality on the basis of the normativity of rationality has thus two, not only one, success condition; that the relevant normative transfer works; but also, and that often goes unnoticed, that the normativity of rationality is accounted for. Nothing comes from nothing.

For sure, there are related projects for which it makes sense to assume such asymmetry. If there are systems running on some norms but not others, it may be nice to show how the norms of the second kind can be incorporated within those of the first. If we are all egoists, then showing that taking morality seriously delivers many personal benefits makes for a neat exercise. But whatever the problem of normativity is, it is not that of somehow bringing people to run by certain norms; it is to show that people should run by certain norms.

My discussion could thus center on any norm; if the normativity of that norm could be successfully explained, the task as I understand it would have been completed. However, because morality and rationality present the two most prominent domains of norms, and because those norms are taken very seriously, they will take center stage in the discussion to come.

1.7. Normativity and Authority

What is authority, then? The functional analysis that took it to be a property of norms helps to roughly locate it. It provides a task analysis. Still, it would be nice to say

more. Calling some norm authoritative is to recommend that norm in some sense; it is to say something on behalf of it; there are, however, many things one can say on behalf of norms; they may be called elegant, to make sense, or to be useful, among other things. We need to distinguish the recommendation of authority in particular from these and others. How to specify the difference?

The task really is to spin an ever wider web of interrelated concepts. In the end we may still not know exactly what we are talking about or that there is something we are talking about in the first place; still, the wider the web, the more connections we can establish, the better. The notion of normativity is in this regard no different than other philosophically tough ones. Consider metaphysical necessity. Most seek to capture that notion by relating it to two other forms of necessity; metaphysical necessity is said to be stronger (more restrictive) than logical necessity but weaker (less restrictive) than nomological necessity; metaphysical necessity allows more than nomological, but less than logical necessity. Analogously, I will try to connect the notion of normativity to with respect to others, some of them similar and some of them interestingly different. I believe that ultimately, though, we have to move away from isolated conceptual webs and turn to substantive theories. We may reach a stage where further concept juggling just doesn't get us anywhere; we then have to think about what normativity could be rather than what *normativity* might mean. We are not there yet, but almost.

1.8. True Norms

Let's turn to truth. Are those norms authoritative that are *true* or *correct* in some sense? Perhaps there is a difference I have not fully appreciated yet between norms merely existing in the background and norms being true. It's not clear, however, what that difference would be. The very idea of true or correct norms edges on confusion; norms do not represent; norms do not have truth-conditions since they do not map what is but what should be. Yet if norms do not even qualify as truth candidates in the first place, it is hard to see how they can be true. The idea must be, then, that true or correct norms are those that exist in special ways, perhaps in ways not dependent on us. Many norms exist because we made them to exist; they may, for instance, be the ones that have *currency* in our society – to use a phrase of David Copp's (1995); they may be the ones that are widely accepted or encoded or enforced or put into law or taken to underlie and guide our practices, or something similar. Yet some norms, perhaps, exist in ways other than that; they – the true ones – may be the ones with authority.

I don't think so; existence, of whatever kind, simply is insufficient for authority. If the idea sounds plausible, it is so only because it has things backwards; the most charitable interpretation is that true norms are those that are authoritative; which, of course, would turn the notion of true norms into a non-starter for coming to grips with authority.

Let's consider a thought experiment to illustrate that even "true existence" and normativity can be dissociated in surprising ways; if, at any rate, *true existence* does not stand short for existence with authority. Suppose there is a single truly existing moral

norm. It is part of the universe in ways that have nothing to do with us. Is that norm thereby authoritative? Not necessarily. Suppose we can find out what the true moral theory is by consulting an imprint in certain stones. For some funny reason or other, the truth about morality is infallibly represented in that particular configuration; just as we are about to finish our decoding, however, we find to our great surprise that the true morality is somewhat different than we thought. While it has a recognizable perfectionist character, it is not quite of the familiar sort; the one and only aim prescribed by the true morality, we find, is to perfect our bodies; athletic considerations alone turn out to be morally relevant – and truly so. Schwarzenegger is a saint, while Gandhi is morally corrupt. Yet after a short period of shock, we feel rather unimpressed by this truth; we do not feel the least bit inclined to comply; we see little reason, for instance, to give up philosophy and head straight to the gym. Now, what are we to think of that indifference on our part towards that truth? Do we need to think of ourselves as behaving as we shouldn't – that we really ought to heed the norm because it truly exists? And if we don't, are we committed to some sort of conceptual/logical mistake? Now, it is obviously true that in displaying such indifference and refusal to comply we violate the precepts of the single true morality; we do not behave as that norm tells us we should; but this constitutes a truism of a rather trivial sort; it is true for any norm that if we fail to comply we do not behave as we should *according to that norm*; yet in rejecting some norm or other we are usually not bothered much that in doing so we self-consciously violating those norms – after all, the authority of those norm is exactly what we reject. Now the difference here is that the norm truly exists; it is not merely of our making. Does that make the difference? I can't think of any good answer why it should. Why not rather

think the universe is messed up? At any rate, it seems perfectly coherent to have this thought; to think, that is, that a norm has no authority and yet that it does truly exist; if that is correct, if truth and normativity can be conceived separate, this would strongly suggest that they are not the same sort of property.

Truth, I conclude, offers little insight into matters normative; at best, true norms are those that have authority; at worst, the idea of true norms is flat-out confused. I believe this also suggests a refreshing vista on traditional ethical debates. The debate between consequentialist and deontologist, for instance, must be about authority or something near enough; it cannot merely be about which sort of norms truly exists – which one really is “out there;” if that was all that was at stake, I believe the point of that debate would be utterly lost; short of some curiosity for funny features of the universe, it is hard to see what practical interest to take in such matters.

1.9. Normativity and Reasons

What about reasons? Normativity can properly be characterized as that which supplies normative or justifying reasons; norms are authoritative if we have adequate reasons to heed them. Reasons talk is now very fashionable and I’m most happy to join in; it is the most convenient normative idiom, and I will use it extensively. I’m not sure, however, how much illumination can be gleaned from such talk. To begin with, we need the very notion of normativity to distinguish normative reasons from other kinds; normative reasons are not considerations we are motivated by, but those we *should* be motivated by, in the *normative* sense of should. But even on a more general level,

normative reasons and *normativity* appear to be interchangeable labels; something is normative if it provides normative reasons, and it provides normative reasons if it is normative. What we want to understand is what it is for a norm to be authoritative, and reasons might offer a clue; yet at the same time, normative reasons seem themselves to emerge from the application to an instance of norms for which the question of authority has been answered; one has reason to do what an authoritative norm tells one to do. Neither notion, then, appears to be primary or to bring much elucidation onto the other.

Still, reasons-talk is most convenient. First, it is pleasantly fine grained; it is more apt to capture all the subtleties pertaining to normativity than other notions such as *ought*. Normativity, they say, is fraught with ought, but *ought* is rather clumsy a notion. *Ought* incorporates an all things considered element which disguises much of what is going on “before” ought-judgments are reached. The normativity of norms need not be reflected in such final ought-judgments. One may regard some norm authoritative even though in certain circumstances one believes one ought not do what it says; one need not thereby deny the norm to carry genuine force. The norm may recommend telling the truth, but now one must not because otherwise people die. Moreover, one need not regard authoritative norms to necessarily bear on what one ought to do either; the norm not to kill may trump the norm to tell the truth in ways that cancel out any possible contribution of the latter to the case; the presence of the truth-telling norm may make no difference whatsoever. It is also consistent to regard norms as authoritative even though one thinks these norms ought not be considered or taken seriously. Consequentialists have demonstrated how norms may be authoritative yet best not be taken notice of in deliberation. Thus, the notion of norms one ought to comply with matches poorly that of

authoritative norms. The notion of reasons, in contrast, is better situated to articulate subtleties such as these and others. It presents little trouble appreciating how what one has most reasons to do one may also have some reasons not to do; or that one has reasons not to take seriously certain reasons one has; we may, for instance, have reasons to reach happiness, but best achieve that goal by not considering these reasons.¹ The expressive power of reasons-talk, then, far exceeds that of ought-talk, and thus is preferable.

Secondly, reasons-talk turns the focus on the practically engaged agent, thereby highlighting the arena where authority truly becomes an issue. We can find norms for anything, including artifacts such as toasters (Thompson 2007); non-defective toasters supposedly toast as specified by toaster-norms; but toasters have no reasons; authority is not an issue for them. Artifacts do not face questions of what norms to take seriously as they never have to make choices between complying and not complying with norms. Agents are different; they have to make choices, and they have to take responsibility for them; the authority of norms matters to agents, both in a forward and a backward looking manner. Agents not only have to decide what norms to take seriously, but also what to think of norms that have been violated in the past, either by themselves or others. Normativity not only supplies guidance, but also critical force; it concerns grounds for criticism as well as grounds for choice. We regularly hold each other accountable for failing to conform to certain norms. Yet an agent's non-compliance with some norm can only constitute grounds for substantive criticism if it was incumbent on that agent to

¹ Dancy is thus mistaken when he writes "A reason is a consideration one ought not to ignore, even if things would go better if one did ignore it." (Dancy 2000:29)

comply with that norm in the first place. Thus, a norm's critical force and its authority go hand in hand.

Reasons with their unwavering gaze on the practically engaged agent also warn us of overly global perspectives on normativity. Some have argued, for instance, that the normativity of norms consists in their social usefulness. Global approaches, however, tend to miss the focal point of authority. Norms are authoritative, because it is incumbent upon particular agents in particular situations to take them seriously. This is not to deny that global facts can be relevant; still, they need to be connected to individual agents. This is most apparent when societal and individual appraisal of what is to be done conflict. Game theory offers an abundance of examples where the maximizing choice from the collective viewpoint is diametrically opposed to the maximizing choice from each individual viewpoint. But even more generally, an agent may wonder whether he has reasons to comply with some norm even though fully acknowledging the social benefits if he did; "it would certainly serve the collective," he may think, "but now here I am and not the collective making the choice; and certainly I am no mere servant of the collective." He may well conclude without confusion that he has no reasons to comply. If the case is an especially drastic one – if he is asked to make substantial sacrifices on behalf of society – he may rather wish to dissociate himself from society altogether; the request to justify to him why he needs to make that sacrifice cannot be answered by sheer force; if that is all that pushes him in directions he is not willing to follow, the project of authority has been abandoned. The practically engaged agent, then, is where the buck stops. Consequently, we may just as well focus on reasons for agents; talk of reasons offers a straight route to that which is of interest in normativity.

Let me close this section with a cautionary note, however. As convenient the reasons-idiom may be, its usefulness can also be overstated. Some appear to think that by merely mentioning *reasons* we can answer most questions about normativity. This is puzzling. What we observe, I believe, is a regrettable tendency for how all too swiftly undue confidence in reasons can turn into philosophical complacency. Words have the amazing power of conjuring up magical realities; religion testifies to that. We need to resist the slippery slope from the prevalence of some talk to the impression that we fully understand that talk to finally the conviction that what we are talking about is entirely unproblematic. It strikes me that some hope if we just talk enough about reasons, the question of their existence may go away.

1.10. Naturalism and Normativity

It is time now to turn from conceptual cartography to my substantive enterprise. I claim that normativity does not exist in the world as we know it. The argumentative strategy I pursue will be this. Calling a norm authoritative, I said, was to recommend that norm in some way or another. I consider various ways of recommending norms consistent with naturalism, and conclude that none of them is sufficient for normativity. Naturalism provides insufficient material for reasons. After noticing this shortcoming, I will then go back and see whether any of the recommendations in tune with naturalism may be practically sufficient even though normatively insufficient; a question, I believe, that can be answered in the affirmative.

But what is naturalism anyway? Naturalism, I believe, is poorly characterized in terms of a defining condition or property – that of being natural as opposed to supernatural. Naturalism approaches ontological questions in a strictly a posteriori manner; what the universe is like and what it contains is not an assumption we start with but one we constantly revise in light of our best knowledge. Naturalism presents an open-ended and inherently incomplete doctrine that incorporates a rationale for why it is open-ended and incomplete. Previous generations of materialists would probably not have felt comfortable with countenancing electromagnetical fields; yet we discovered them anyway. The lesson we learn from this is that we are likely in for more than one surprise. As Haldane famously said, “my own suspicion is that the universe is not only queerer than we suppose, but queerer than we *can* suppose.” (Haldane 1927:286) Naturalism construed as an inflexible and principled doctrine would rapidly unnaturalize itself.

Still, I believe it comprises two broad ideas; both have been subjected to innumerable tests, and survived. First, that we live in one unified world where everything is connected; there are no isolated phenomena that are outside the reach of everything else. Secondly, that the world is structured; complex phenomena are systematically related to less complex phenomena. As the physicist Murray Gell-Mann has put it, “you don’t need more to get more.” (2007) The second idea implies a weak form of reductionism; nothing that exists comes from nowhere. This articulates a belief in the fundamental intelligibility of the universe, even if it may be beyond our grasp; complex phenomena could in principle be explained by the complex interaction of less complex phenomena. This form of reductionism is weak, and is a far cry from various stronger straw man versions that have been successfully refuted, such as the notion that higher

order laws can semantically be captured by lower-order laws in conjunction with bridge laws. It is surprising that many philosophers still hold on an outdated paradigm of reductionism, without taking seriously the idea that the common conviction that nothing comes from nothing requires some sort of account. (Kim 1998)

Naturalism constrains the search for the normative in the following manner. First, to repeat, the constraints it suggests are non-final and subject to revision. When I claim that naturalism provides insufficient material for reasons, what I am claiming is that this is so only in light of our currently best knowledge of what there is. Secondly, naturalism rejects primitive normativity as an unstructured phenomenon unconnected to everything else. If normativity exists, we must be able in principle to situate it in the interactive complex we call our universe; call this the principle of accommodation. We must also be able in principle to explain normativity in virtue of some aspects or other of that world; that is what it means that we cannot take it as a given or ontologically primitive; call this the principle of explanation. I think that those two ideas broadly capture what most would take away from naturalism. They are particularly well suited for phenomena we have no immediate acquaintance with, and certainly normativity is such a phenomenon – we cannot simply point towards it. If someone proclaimed to just have seen or felt or smelled normativity we would likely take him to misunderstand what normativity is. *Normativity*, thus, behaves somewhat like a theoretical term; it purports to pick out a phenomenon the existence of which needs to be established indirectly. To explain it and to show it existent are intimately related projects.

Non-naturalistic accounts of normativity have seen an astonishing comeback recently. Non-naturalists believe normativity is an ineliminable and primitive part of our universe; if we cannot explain it, we have to accept our inability to situate normativity within the world as we know it. In *Moral Realism – A Defense* Schafer-Landau writes (2003:205):

Those who affirm [that there are no intrinsic reason-giving facts] will say that all reasons derive from an agent's perspective. For consider the alternative: if reasons exist regardless of one's beliefs, desires and interests, then where do they come from? ... To insist that a set of facts could contain within themselves normative authority for agents, regardless of their outlook on life, seems obscurantist, and gives the appearance of prematurely cutting off any helpful explanation of normativity. If this is obscurantist, I think we have no choice but to embrace the mysteries. I think that intrinsic normativity is ineliminable.

My argument is channeled to those who like me do not want to embrace the mysteries. Still, I consider non-naturalists my greatest ally, and here is why. I concur that important parts of normative discourse are indeed currently committed to normativity; that's why I am an error theorist and not an expressivist, say. But most importantly, I share the same skepticism that normativity cannot be accommodated in the natural world. Their arguments to that effect thus offer invaluable sources of inspiration.

Chapter 2

Against Non-Reductive Normative Realism

2.1. Introduction

This chapter wrestles with non-naturalistic or non-reductive normative realism, a position that continues to gather widespread and prominent support. The view holds, in David Enoch's succinct summary (2007:21), that *there are response-independent, non-natural, irreducibly normative truths, perfectly objective and universal ones, that when successful in our normative inquiries we discover rather than create or construct*. This view – conveniently named *realism* within the chapter's confines – represents not only the sharpest adversary to my own position, but also levels the perhaps most fundamental challenge to it. It is proper, then, to address realism and its challenge first before I turn to naturalistic or reductive normative realism in the following chapter.

Realism confronts anti-realism in stark transcendental fashion by questioning whether the latter's global denial of normativity represents a position that can be coherently adopted and defended. There are no reasons in the fundamental normative sense, I claim, reasons that would provide authority for norms, standards or rules. Yet settling on what to do and accept appear reason-guided enterprises, and settling on whether to accept global normative anti-realism is one such enterprise. Surely enough, then, the anti-realist cannot coherently support his view by citing reasons in its favor after he just rejected them throughout. In recommending his view, what is he doing, then? Will he not have to become silent, ushering himself off the philosophical scene? Worse, a defense of his position may even testify to its refutation, since it exemplifies what it supposes impossible: providing reasons for accepting a hypothesis. In response, I show how to support anti-realism without deserting its tenets. I begin with the admission that making the case for and against philosophical hypotheses essentially are norm-driven enterprise. We need norms supplying criteria of argumentative correctness and success. Yet we need not presume the relevant norms underwriting philosophical argumentation equipped with practice-external authority; our employment of them is not bound up with and contingent upon the metaphysics of realism.

Despite their sharp contrast, realism and anti-realism have actually a great deal in common as well. They share a similar conception of what is at stake; they fundamentally agree on what normativity is. This by itself turns them into allies in a remarkable number of ways. For instance, realists and anti-realists look eye to eye when it comes to the plentiful intermediate attempts to explain or reduce normativity. They converge in their assessment that such attempts are unviable for they either fail to explain normativity or

simply lapse back into realism; that there cannot be normative engines that successfully explain authority without already presupposing it. The principle both accept is no normativity in, no normativity out, or conversely, that by denying the irreducible normativity of some phenomena we deprive it from any recognizable form of normativity altogether. “Moral and other evaluative facts have a feature that no natural fact could have,” writes Jonathan Dancy “namely, normativity. If we try to identify moral facts, or facts about what we have most reason to do, with natural facts, their normativity is lost.” (2006:132) This shared belief places realism and anti-realism in a dialectical situation similar to that between libertarians and hard determinists in the context of the free will debate. Libertarians and hard determinists too share a common mistrust concerning the various “softies” in between – suspecting that the compatibilist majority merely seeks a way out by conveniently changing the subject. Thus, however profoundly realists and anti-realists disagree on a variety of issues, they must at least grant each other to have gotten the subject matter right. All the more reasons, then, to start with realism. It encompasses an understanding of normativity that proves most helpful for clarifying what the debate is all about, and provides the perfect platform for formulating and defending my global denial of authority.

Here, then, is what I shall to do in the chapter. I will first provide a systematic and sympathetic reconstruction of realism that incorporates a variety of recent and prominent articulations of the view. Realism, I believe, is what most philosophers in ethics are in fact committed to – even though the commitment is not always explicitly acknowledged and on occasion even ridiculed. This is a mistake the nature of which will hopefully emerge from the analysis I offer. Realism is an internally sound and well thought through

position that has proven resistant in the face of considerable counterattacks. In the end, though, I do reject realism, and so I'd like to say why and why not. After that, I turn to the heart of chapter, the refutation of the transcendental argument.

2.2. Normative Non-Naturalism

What does normative realism assert, and what in response does global anti-realism deny? Realists obviously affirm the existence of some normative reality. What is distinctive of their view is of what kind they take that reality to be. The label non-naturalism is intended to denote a certain way in which the normative is presumed independent of the natural. A recent statement of the view has it that “there are ethical facts that obtain independently of our actual ethical beliefs or attitudes or practices, both on an individual and a societal level, *at least* in the sense that such facts about what is right, or what is good for us, or what reasons exist are not simply a direct function of these things as they stand.” (FitzPatrick 2008:162) Realists on occasion prefer using the label *primitive* instead of *non-natural*, but what does *primitive* amount to if not an ontological addition to what we already get in terms of and in virtue of the plain natural? Sorting out in what ways the normative is primitive or independent from the natural presents a subtle matter. It is especially easy to misunderstand what is and is not supposed non-natural. What will be required is a careful and systematic exposition of how the normative relates to the non-normative according to realism.

Let me first, however, comment very briefly on the contrast between the natural and non-natural itself. This is not an essay in general metaphysics, and I'd rather not get

entangled too deeply with the difficult question of how to precisify this distinction. I am unaware of any straightforward and uncontroversial approach to use as a basis for what follows. Still, I believe the situation is actually not as conceptually dire as it may initially appear. However we may ultimately choose to fill out the exact metaphysical details of the natural/non-natural distinction, a corresponding position of normative non-naturalism should readily emerge. In the context of the present debate, then, we shall be content with having the terms *natural* and *non-natural* figure as placeholders, and regard my occasional substantive remarks as primarily serving expository purposes.

Let me start out with some recent statements by three prominent realists, first by Derek Parfit, then Schafer-Landau, and finally Thomas Scanlon, all of whom I propose to quote at considerable length. In the section two of his excellent essay *Normativity*, Parfit offers a most clear-headed characterization of realism that contains all the important labels, a position he then proceeds to defend in depth. “Irreducible normative truths” he writes “are most unusual” for “it is not obvious how such truths fit into a scientific world-view. They are not empirically testable, or explicable by natural laws.” He then raises the question “If such truths are not empirical, or about features of the natural world, how do we ever come to understand them?” In response, he concedes “non-reductive realists ... do not give helpful answers to these questions. ... we can explain some normative concepts, but only by appealing to others.” In sum, “normative concepts cannot be explained in non-normative terms. Nor can we say much to explain how we understand these concepts, or how we recognize normative truths.” However, we need to keep in mind that “if there are normative truths, these are of a distinctive kind, which we should not expect to be like ordinary, natural truths.” We can find, he believes, some helpful

analogies in modal concepts such as *possible* and *necessary*, which also “cannot be explained in empirical terms ... and are not made true by natural laws.” Parfit then briefly turns to “some writers who try to explain what normativity is,” an undertaking “that cannot be helpfully done.” Taking the *via negativa*, he goes on “though we cannot explain what normativity is, or what normative concepts mean, we can say what normativity is *not*, and what these concepts do not mean. It could not be true that, as naturalists claim, normative statements mean the same as, or report the same facts as, statements about natural facts. Naturalists ... mistakenly conflate these [normative] truths with the natural facts which, according to these truths, have normative importance.” As we shall see momentarily, especially this last statement proves most essential.

In *Moral Realism – A Defense* Shafer-Landau provides a remarkably well-written and very systematic defense of realism, in particular of non-natural moral realism. He considers the essential feature of realism to be “its endorsement of *stance-independence*.” “Realists believe,” Shafer-Landau writes, “that there are moral truths that obtain independently of any preferred perspective, in the sense that the moral standards that fix the moral facts are not made true by virtue of their ratification from within any given actual or hypothetical perspective.” Shafer-Landau considers realism “a view about the status of ... normative theories [that] insists that the truth of any first-order normative standard is not a function of what anyone happens to think of it. Such standards, if true, are not made true, and, in particular, are not correct in virtue of being vindicated by some process of (inter)personal election or approbation.” (15-16) And later on he confronts a challenge to the very aspect of anti-reductivism in an intriguing way, starting with a paraphrase of the charge: “Those who affirm [that there are no intrinsic

reason-giving facts] will say that all reasons derive from an agent's perspective. For consider the alternative: if reasons exist regardless of one's beliefs, desires and interests, then where do they come from? ... To insist that a set of facts could contain within themselves normative authority for agents, regardless of their outlook on life, seems obscurantist, and gives the appearance of prematurely cutting off any helpful explanation of normativity. If this is obscurantist, I think we have no choice but to embrace the mysteries. I think that intrinsic normativity is ineliminable. (p. 205)

In his 2009 John Locke Lecture on normativity, Scanlon says (8.27) "I also maintain that truths about reasons are irreducible normative truths. Not reducible to, or identifiable with, truths of other kinds, such as truth about the natural world of physical objects, causes and effects. So I am what might be called a reasons-fundamentalist." He then asks about the grounds of normative truths (10.36): "In virtue of what are claims about reasons true when they are true? In particular, does the idea that claims about reasons can be true or false independent of our options about them and that truths about reasons are irreducibly normative have unacceptable metaphysical implications? ... How are facts about reasons if there are such facts related to natural facts? They are not entailed by natural facts all agree, but they tend not to vary unless natural facts vary, and this strange relation of non-entailment that some kind of rigid binding may seem puzzling and in need of explanation."

With these statements in hand we can now turn to a systematic reconstruction of realism. What is most important to get clear on what is and is not supposed non-natural according to realism. The normative obviously engages the natural world as it is; all that

is relevant must be built up entirely from natural bits and pieces – what else is there to work with? Realism does not contest that. The role it assigns to the non-natural has another source altogether. As I understand the view, what is supposed non-natural is the *determination of what is relevant* rather than *that which is relevant*. This is what I take Parfit to mean when, in the above quotation, he urges us not to “conflate these [normative] truths with the natural facts which, according to these truths, have normative importance.” Realism starts from the recognition that there are multiple logical alternatives for what has normative relevance all of which equally consistent with a thoroughgoing conception of the natural world. None will be simply ruled out on natural or conceptual grounds alone. Realism is the hypothesis, then, that there are normative truths concerning what is relevant, even though those truths are not determined by the natural or conceptual. There are, first, all the natural facts that constitute the world as it is; normative truths then simply concern which of those natural facts are normatively relevant, according to realism.

The crucial distinction to notice, in Jonathan Dancy’s words, is “the difference ... between facts that might be mentioned in answers to the question what to do, and the facts that those facts are relevant to the question what to do. To give a very contentious example of this difference: the fact that this action would make many more comfortable and none less comfortable could well be mentioned in an answer to the question whether it is the thing to do, but is not the same fact as the fact that it is relevant to what to do.” (2006:137). Normative facts, in this picture, figure as meta-facts; facts regarding which regular natural facts have normative significance. As Dancy continues “It is these metafacts that I think of as the central normative facts, by reference to which the

normativity of all others is to be understood. Each such fact is the fact that some other fact stands in a certain normative relation to an action (or a belief or a feeling or a desire ...).” It is easy and common to mislocate the central normative facts Dancy is alluding to at the wrong level. This is exemplified, for instance, by Jackson’s caricature of realism, a position he takes to imply that “someone who says, ‘I see this action will kill many and save no one, but that is not enough to justify my not doing it; what really matters is that the action has an extra property that only ethical terms are suited to pick out’.” (Jackson, 1998, 127-8). Yet normative facts, as the realist envisions them, do not posit any extra properties to be adjoined to the plain natural; rather, they take the natural as it is and dignify it with normative significance. To return to Jackson’s example, the fact that an action will kill many and save no one is indeed all we need. The relevant normative truth is simply that this is normatively relevant (that it takes the action off the table of permissible options, or that it provides reasons not to do it, or something else). It is potentially misleading, then, to read realism as postulating extra normative properties of any kind. All it does is to identify which natural configurations are relevant.

The underlying idea can be best brought out in terms of the notion of normative functions or alternate mappings from the natural onto normative status. I will illustrate my own systematic reconstruction of realism by going through three prominent normative notions, namely *good* and *bad*, *right* and *wrong*, and *normative reasons*. It is my belief that these present three distinct normative domains, where the content of each cannot be adequately characterized in terms of the others – a claim I shall defend in later chapters. For now, and in light of that conviction, I’d like to separately specify for each domain what in my understanding realism amounts to. I need to stress right away,

however, that I am presently concerned only with providing a working characterization of non-naturalistic realism, and thus my focus will be schematic and rest only with a few relevant details; the following is certainly not intended as a comprehensive analysis of the respective normative notions.

Start with *good* and *bad* or *value*, then. Different theories of value identify different features as valuable: pleasure, knowledge, autonomous choice, perfection, harmony, unity in multiplicity, preference-satisfaction, etc. Most if not all of the features singled out as valuable by some theory or other are clearly part of the natural world. What proponents of rivaling views of value contest is usually not that the features their opponents advance as valuable do exist. Few adversaries of hedonism, for instance, deny the existence of pleasure. They take issue only with the normative status hedonists assign to it, namely that of being singularly valuable. Each position on value, then, could be summed up in terms of a complex function that assigns to each value-candidate one of three numbers: *one* for valuable, *negative one* for disvaluable, and *zero* for value-neutral. Returning to the list from above in the same order, hedonism would correspond to the function (1,0,0,0, ...) assigning the negative -1 to pain only. *Intellectualism* would be the alternative view that corresponds to (0,1,0,0,0,...), Nozick's late rendition of some Leibnizian ideas in terms of organic unities or of unity in multiplicity would perhaps correspond to (0,0,0,1,1,1,0,0,0,...), *optimism* to (1,1,1,...) and *pessimism* to (0,0,0,...). Given some basic mathematical combinatorial facts, there will be lots of alternatives for sure. With respect to each alternative it will be the case, then, that what it postulates as valuable are just plain natural features. Value will readily supervene on the natural; the particularly valuable will perhaps be fully constituted by the natural: According to

hedonism, there is nothing more to value than plain natural pleasure, according to intellectualism there is nothing more to value than plain knowledge, and so on. Virtually all theories of value are naturalistic in the sense that what they identify as valuable are features entirely natural. However, it is also true that virtually all theories of value are non-naturalistic in the sense that the determination of what is valuable is not supposed naturalistic. Even a comprehensive conception of the natural would leave it open which function gets it right. What realism asserts is that there is a truth of the matter of which theory of value is correct, and that this truth is not of the natural sort.

The same goes for *right and wrong*. It is entirely compatible with realism that all candidate ethical theories can be taken to share the presupposition that what is morally relevant must be natural. No moral theory, according to realism, needs to make the implausible assumption that what matters morally is something other than plain natural configurations. Again, the issue is not whether morally relevant features are or are not natural. Realism is not forced to accept, and is even free to reject, this second disjunct. Rather, the question is whether the fact that this rather than that feature or configuration is morally relevant is a natural fact; whether we could adjudicate between alternative mappings of natural configurations onto moral status solely in virtue of natural observations. Realists grant we cannot, and yet assert that there are moral truths; that some mappings but not others are in fact correct.

Focus, for convenience, on the narrow set of the traditional ethical candidates utilitarianism, deontology, and virtue-ethics. Each of the three theories advocates alternative mappings between natural features and moral status. Utilitarianism maps

moral status onto behavioral options in terms of their agglomerative effects on overall pleasure and pain; deontology does the same in terms of what maxims can be universalized, and virtue-ethics in terms of whether beneficial character-traits are exemplified. For short, utilitarianism (U) maps moral status onto u-features, deontology (D) onto d-features, and virtue-based ethics (V) onto v-features. The important point is that neither U nor D nor V need to think of the respective features they identify as morally significant – u and d and v – as non-natural. Moreover, all sides may grant the same for their competitors, namely that no theory needs to postulate any funny non-natural creatures to carry moral status. For the sake of the present argument we can simply accept that effects on pleasure and pain (u-features), maxims (d-features), as well as character-traits (v-features) are all part of the natural order. Dissenters of utilitarianism, for instance, do not usually deny that actions affect future distributions of pleasure and pain. What utilitarianism's adversaries contest is an altogether different matter: namely that this kind of differential impact is the sole carrier of moral relevance. It would be entirely futile and wrong-headed if in response to their opposition utilitarians simply reaffirmed that behavior does have causal effects. For utilitarianism to emerge victorious, more is required than this plain truth; in particular, what is required is that their preferred mapping, rather than that of their challengers, turns out correct, even though none appears to be ruled out on natural grounds alone.

It presents no major obstacle, then, to see how each moral theory, if correct, could make good on the idea that the moral status of particular actions is comprehensibly constituted by their natural features; the concrete moral status of an action would not only weakly supervene on, but even be determined and fixed by natural facts. For utilitarians,

being morally superior just is resulting in better effects; for deontologists it just is being chosen in accordance with (or as a result of the recognition of) the relevantly universalizable maxims; for virtue-ethicists it just is exemplifying beneficial character traits. There is no need for any additional, potentially non-natural components to complete the picture. Indeed, the ease with which each theory can proclaim constitutive relationships between the natural and the moral has tempted some to prematurely conclude that theirs is a naturalistic moral theory. Not so fast says the realist, and rightly so. The important qualification, of course, is *if correct*. Each U and D and V offer rivaling identifications of moral correctness; the question is whether any of them is correct in that identification. Suppose one does get it right. Is that a fact that is fixed and determined by the natural order of things, and could we in principle ascertain and settle its presence solely in virtue of some natural method of investigation? According to realism, we presently have no good evidence to believe so, and a commitment to moral truth thus translates into an ontological commitment to certain truths of the non-natural sort. Since all sides may very well be in agreement over all natural facts while remain in disagreement over moral relevance, their disagreement consequently does not concern the natural but the non-natural.

Finally, turn to normative reasons. It is now customary to consider *normative reason* the quintessential normative notion, perhaps even the mortar that infuses all others with their normativity as well. “All normative phenomena,” Joseph Raz opens a recent address, “are normative in as much as, and because, they provide reasons or are partly constituted by reasons.” (2010:5) This is a mistake I must wait to examine later. For now, and going with the flow, the idea of a normative reason, as I argued in the first chapter,

incorporates norms presumed authoritative. Norms themselves, to continue the working suggestion of the previous paragraphs, can be seen as complex functions, mapping as appropriate certain initial conditions onto certain responses. Traffic norms, to pick an instance, specify which types of automotive behavior constitute appropriate responses to traffic-related situations. Reasons then come into play once certain norms and their corresponding mappings have become imbued with authority; or, conversely, norms acquire authority once they issue an appropriate provision of reasons. To the extent to which traffic norms prove authoritative for us – possibly through the interaction of more general norms – we will have reasons to act accordingly.

What is important is not to presume one of the notions *reason* and *authority* prior and capable of explaining the other, but to learn something by noticing their interrelatedness. Reasons are considered favorers, period, not just favorers-according-to-some-norm-XYZ. This initial dependency of the favoring-relationship upon some specific norm is supposed dissolved once the relevant norm itself becomes authoritative, period. Normative reasons, then, incorporate “correct” and “authoritative” mappings between situations, agents, and responses. Saying that someone has a reason to help the child in need while passing by amounts to saying that given the situation in which he finds himself helping is the thing to do or is the appropriate response to take; not just according-to-some-convention-or-norm, but simpliciter. Having this reason supposes, then, that the mapping child-in-need onto helping is correct whereas that of mapping child-in-need onto passing-by-without-helping is incorrect. The relevant question is what renders one mapping correct whereas the other incorrect? For each combination of situations, agents, and responses there is a virtually infinite number of such alternative

mappings. Realists believe that there are correct mappings and that which are is neither entailed by, fixed or determined by natural facts, nor can be established solely by natural methods. What realism assumes is that there is a truth of the matter of *what norms are authoritative* or of *what authoritatively favors what*, and that this truth is again not of the natural sort; which is not to deny that *that which favors* and *that which is favored* are entirely natural configurations.

We must then again be clear about what is and is not presumed non-natural here. The particular states of affairs that carry normative significance according to the relevant authoritative norms are usually just plain natural configurations. *What was the reason to help?*, we ask, to which we answer *that the child was in need* – a response that points at nothing more than a plain natural configuration. Yet this response remains somewhat elliptical. For sure, that which is reason-related to our helping response is not itself a non-natural phenomenon. It solely consists in the child's occupying a certain spatial-temporal location being quickly filled by water. What we identify as reasons will quite generally be plain natural phenomena – what else could it be? Yet what is non-natural, according to realism, are not the reasons-relata, but the normative relation of *favoring-with-authority* itself. "A person who accepts [that X is a reason for doing A] takes a certain belief to be warranted," Scanlon correctly notices, "namely the belief that the relation 'counting in favor of' holds between X and doing A. That this relation holds will then be a 'non-natural' fact, that is to say, a fact that is neither merely a fact about our psychology nor an ordinary empirical fact about the world outside us." (1998:58) The underlying idea is again that a complete specification in natural terms of the child's situation would not entail which response to take, or to which particular response it is reasons-related. That

this is not imminently obvious has mostly to do with the fact that our descriptions of situations usually proceed in already value-laden and reason-apt terms. *In need of help* obviously suggests *helping* as the course to take. Yet the callously indifferent person need not miss a natural detail in that situation. What renders correct the mapping child-in-need onto helping whereas incorrect the mapping child-in-need onto passing-by-without-helping is not a natural fact, even though it is a fact, or so the realist believes.

My characterization of realism thus far did proceed in somewhat abstract and structural terms. I emphasized solely that realists take normative truths to have some sort of independence from the natural, without specifying what sorts of things in particular realists take the normative to be independent of. Yet since their naturalistic opponents usually identify the normative with a confined set of prominent candidates, I shall explicitly name what realists think normativity is not. First, realists take normative facts to be evidence transcendent in the sense that they obtain independently from what we taken them to be. They are not projected, or constructed, or emerge as the shadows of our attitudes and beliefs in any other way. Realism believes in a profound difference between being correct and being found correct, or any other function from what we agree upon or endorse as correct. When we *truly and responsibly* believe what is valuable, right, or favored, we do so in virtue of normative facts, and not vice versa. The exact nature of evidence-transcendence is a complicated matter. For our purposes, it shall suffice to think of it in similar ways as realists do in other domains. The correctness of our conception of the normative world depends on the normative world, not the other way around, just as the correctness of our conception of the external world depends on the external world, and not the other way around.

Secondly, realists take normative truths to be independent from motivational and other dispositional psychological facts, be it of an individual or societal nature. Realists reject all constructivist approaches, who, despite all their variations, always offer essentially the same answer in response to the question of the Sphinx: it is man, or it is us and our consent, or it is society who sets up which normative mappings are correct and operative. Not so says the realist. That suffering provides reasons for relief is not to be explained in terms of what we are motivated to do in the face of suffering, even after reflection, or what reactions we are disposed to cherish as a group, etc. The common equation of normativity and motivation is what non-reductive realists in fact scorn most. They correctly urge us to keep distinct normative force from motivational force. This goes in both directions. As mentioned, having a reason to ϕ does not consist in being motivated to ϕ under some non-trivial and non-normatively specified conditions C. In turn, having a reason to ϕ does not require being motivated to ϕ under conditions C. For sure, tinkering with C might very well generate a true bi-conditional between the normative and the motivational. This may be most easily achieved by having C to include normative qualifications such as recognition of what reasons pertain and flawless execution of practical rationality; which, obviously, would not get us anywhere. Upon close inspection most biconditionals that have been suggested to bridge the normative and motivational contain normative vocabulary in the definiens part of, if only in the suggestive and intolerably vague phrase “*in the right way*.” Motivation need not be normative and normativity not motivational, even if both can be linked bi-conditionally under some trivial conditions C.

2.3. What to Make of Realism – Impotent Metaphysical Arguments

Realism raises a variety of metaphysical, epistemological and practical issues. Let me briefly outline my own assessment. Ultimately, the case for or against realism must prove inconclusive, for, I suppose, structural rather than purely evidential reasons. It is the very manner realism is set up, I believe, that preempts arguments that would depart solely from premises judged acceptable or even tolerable by proponents and opponents alike and that would conclude in a judgment about the truth of realism. This is not universally recognized. A good number of arguments advanced in the literature portray little but an unfortunate misunderstanding of what is at stake. In recognizing why they fail, indeed must fail, we may actually learn something about what position realism is. For the sake of intellectual honesty it is important, then, to acknowledge the limitations of what we can argumentatively achieve. But this has a flipside as well. Once this recognition has been explicitly stated, it becomes rather boring and repetitive to point out time and again why certain challenges are inconclusive. The task at hand will be to distinguish the more promising candidates among the pool of all inconclusive arguments, in the expectation of getting a dialogue going beyond the observation that nothing will ever be decisively disproven.

Let me begin with the metaphysics of realism. Realism is first and foremost an ontological thesis; it would be desirable if its truth could be assessed accordingly in direct ontological fashion. It remains unclear, however, whether this can be done, and whether the best we can do in the end may not be to resort to highly indirect ways of assessment. For now, consider the most prominent direct argument against realism, which proceeds

from the purported tension between realism's ontology and a thoroughgoing natural conception of the world we live in. The nature of the tension is uncertain, however. Realists stress time and again the compatibility of their view with any such thoroughgoing natural conception of our world. There is nothing paradoxical about this. Realism simply does not issue any empirical implications which in turn could stand in competition with a natural conception in the first place. As the slogan goes, realism is not in the business of explanation but of justification. In this respect it is quite unlike any strand of supernaturalism which does aspire to make explanatory contributions for natural events; when, for instance, the sickness of a child is accounted for by some evil stroke of a witch. Realism, in contrast, does not cite non-natural facts for the sake of the explanation of natural facts. The non-natural reality it postulates does not interfere with the natural reality in ways that could put it at odds with any comprehensible natural conception. Realism in this sense corresponds to deism in theology, a view that became increasingly popular with the triumphal procession of natural science. Realism and deism alike have a design that immediately immunizes them from any conceivable empirical criticism. Since realism is empirically unfalsifiable, the question of whether to accept its non-natural ontological addition to the natural cannot be directly adjudicated on the basis of our best knowledge of the natural world.

This bears considerable dialectical significance. Some have rejected realism solely on grounds of its non-natural metaphysics. But notice that the truth of naturalism of the sort required for such a refutation is itself what is at issue and moreover cannot be determined by appeal to science or natural method alone. The acceptance and the denial of naturalism alike lack any empirical footing. As a matter of sheer logic, the hypothesis

that there is nothing beyond the empirically accessible is not empirically accessible. Empirical science deals only with the empirical; whether there is an extra-empirical reality is beyond the scope of the empirical. To use a neat picture of Wittgenstein's, one cannot determine what, if anything, is outside the circle by staying within. Neither realism nor anti-realism can hence fully escape a certain residue of dogmatism.

What this makes readily apparent, though, is why various arguments against realism must fail for their misconception of what realism involves. Consider, for instance, versions of the argument from explanation. Its core idea is to make a certain kind of explanatory indispensability a criterion for ontological inclusion. Since normative facts plainly are explanatory dispensable (*hallo!*), their ontological standing must become so as well, or so the argument claims. Yet this argument flat-out ignores that realism was never purporting to have any explanatory impact in the first place; it merely claims that in addition to explanatory facts that it has no business with there are justificatory facts as well. To insist that all facts must be of the explanatory rather than the justificatory sort is simply to beg all relevant questions.

This has an interesting flipside. Some philosophers have sought to substantiate the presence of a normative or moral reality by allegedly spotting an explanatory contribution that reality makes. The wrongness of slavery is supposed to explain why it ultimately became abolished. Call this the positive argument from explanation. Now this argument gives rise to a funny set of issues I cannot even begin to tackle in full here – I will come back to some of those issues in the next chapter. But notice for the moment that there are different ways in which the normative could become ingrained in some explanatory

enterprise, not all of which bear dialectical potency. In a first manner, one may hold that the natural configurations singled out as normatively significant play some explanatory role. This, however, turns out such an uncontroversial contention that it cannot argumentatively differentiate between the competing players in the meta-normative debate, since everyone in principle may accept it, including realists, anti-realists, constructivists, relativists, etc. It was antecedently agreed upon, I presumed in my analysis, that whatever is morally relevant must be plain natural configurations of the world we live in, and hence will have causal ramifications. Pleasure is causally interconnected with other natural phenomena, and if hedonism is correct, then that which is valuable obviously plays some causal or explanatory role. No problem. Moreover, all participants can even accept, in principle, that normative or moral categories render particularly salient certain natural phenomena and their interrelations; that they afford certain sorts of insight not to come by in alternative terms. Why did John and Paul both end up in prison? Well, they both must have done something wrong – and this may be the only commonality between John's and Paul's behavior. The availability of such explanations is actually most easily accounted for by constructivists; for them, moral categories reflect our collective attitudes in some way or other. No big surprise, then, that moral terminology can illuminate various phenomena, including our moral reactions! Yet realists worth the label must surely reject constructivism. An argument that does not favor realism over constructivism is no good as an argument for realism.

So if the positive argument from explanation is to get some dialectical millage, it must identify another way in which the normative supposedly impacts the natural. And that is not easy. The most popular proposal ties explanatory contribution to the truth of

certain counterfactuals. That, for example, had slavery not been wrong, it would not have been abolished. Yet that does not help, even if the latter statement proved correct. The truth of counterfactuals is too easily secured, and does not require any causal or explanatory connection between its components. This is immediately apparent from phenomena that give rise to stable epiphenomenal characteristics, such as the shades of the parts of a moving carriage. It appears as though the shade of the horses pulls the shade of the wagon; but this is obviously not so, despite the truth of this counterfactual: Had the shades of the horses not pulled in this direction, the shade of the wagon would not have moved accordingly either. This is true, because in the nearest possible world where the horse-shades are arranged differently the real horses occupy different positions as well, which then affects the position of the wagon, which in turn alters the wagon's shade. The only causal and explanatory connection is that between the real horses and the real wagon, and it only so happens that their shades reliably trace that connection. The same is true in the normative and moral domain. Whichever ethical theory emerges victorious, it will be true that if slavery had not been wrong, that could have been so only because slavery would have had alternate natural properties in that scenario (such as causing fun instead of misery), and that obviously changes the entire game. Since it is agreed by all sides that the natural features of a particular event fix its moral status, difference in moral status implies difference in natural status, which in turn is all the difference explanatorily required. Even Descartes, the textbook dualist, holds that there is a relationship between the mental and the physical that supports counterfactual statements – due to stable correlations between the mental and the physical, potentially established and secured by God. But the mental obviously does not explain anything

physical in that picture. So the truth of certain normative-factual counterfactuals won't secure a positive explanatory contribution of the normative either.

The only way I can see the normative impacting the non-normative in a meaningful manner is this: Suppose, at time t_1 , hedonism is correct. Value-status is attached to pleasure. Now change the relevant normative switches while leaving everything else equal. Now, at time t_2 , value-status is attached to preference-satisfaction. Would that switch have any causal impact? This is a funny scenario, for sure, potentially metaphysically impossible, even if logically possible. The answer to that question is not entirely clear – to the extent to which the question is sensible in the first place. I'd say the answer is no, and so does the realist. In any case, since neither of us takes the normative to be in the business of explanation but rather in that of justification, we find the positive argument from explanation off target from the very start.

Realism, I thus conclude, is neither empirically assailable nor supportable. Is this good news or bad news for the view? Neither, I'd say. It presents no step forward because it does not suffice to merely prove ones opponents' charges ineffective, and yet that is often all realists do, even by their own admission. Parfit, for instance, writes: "If we believe in irreducibly normative truths, we are what Korsgaard calls dogmatic rationalists. As Korsgaard notes, since these rationalists have little positive to say, they are 'primarily polemical writers', who explain and defend their views by attacking other views. That is what, in this essay, I shall mostly do." (2006:332) Refuted criticism does not increase probative force. It presents no step backward either, despite the verdict of some traditional empiricists who consider unfalsifiability the ultimate sin, worse even

than being false. Sentiments have tempered down since logical positivism and Popper. Opposition to empiricism is again a reputable option, let alone intelligible and worth consideration. And even the brands of empiricism currently fashionable are more moderate. The empiricism I favor would certainly recommend abstaining from positively accepting any hypothesis that lacks empirical support, but not necessarily going as far as recommending flat-out denial instead of agnosticism. Of course, if there is a presumption in favor of a positive case being made before the need arises for demanding sound refutations – a presumption that is certainly part of scientific practices – realism potentially is in trouble; in trouble, that is, as long as it cannot identify other sources that lend positive evidence in its favor. And to those other sources we must ultimately turn, in both the positive as well as the negative case. The best we can do, in my assessment, is to discuss more indirect motivations for and against realism. In the absence of opportunities for face-to-face combat, we must resolve to focus on the supply lines instead. In the end one may simply believe realism, or one may not.

One indirect argument we can put aside quickly. This is the familiar *partners-in-crime* move realists on occasion advance. Realism is supposedly alike other hypotheses postulating non-natural realities, most prominently ones advanced in the philosophy of mathematics. And that, it is hoped, takes away some of the pressure resting on realism. Yet either the argument clumsily demonstrates what we already know – that an insistence on the truth of naturalism in order to refute realism amounts to mere question-begging – or it is simply no good. An initial suspicion already arises from the very labeling “partners in crime.” I always thought crime is a bad thing, and how exactly the mere presence of accessories could offer any legitimate excuse or justification is beyond me.

But more to the point, the argument has two serious shortcomings. First, normative-truths have almost nothing in common with the usual accessory, mathematical truths, apart from the hypothesis that both obtain non-naturally. Normative truths and mathematical truths supposedly have the same modal status, obtaining necessarily. But even that is not so. At best, normative truths are only metaphysically necessary, while mathematical truths are logically necessary – a huge difference. Moreover, whether normative truths are or must be metaphysically necessary is also unclear. This is certainly often dogmatically affirmed, but rarely defended. That it could not be the case that consequentialism is true in one world while deontology in another requires some argument. Certain strands of constructivism and relativism would clearly deny it – without apparent conceptual confusions – and suppose it possible, perhaps even likely, that alternative moral codes will be operative in different worlds. Notice these modal scenarios are consistent with world-by-world, even though not with across-world, supervenience. But only the former appears supported by actual moral practices and observations, and has potential significance for us, as when the injunction ‘treat like cases alike’ provides the basis for principled criticism. Yet in any case, mathematics and normativity are utterly unlike in a number of important ways that undermines any probative transfer. Most pertinently, mathematics, unlike normativity, is non-normative (sic.), part of a well-developed formal discipline, successfully and essentially employed by science, departing from uncontested axioms according to formal procedures, etc. On reflection, the alleged analogy between mathematics and the normative actually turns out quite a major disanalogy.

As a second point, if we take the partners-in-crime move seriously, there simply will be too many to swallow. One partner is theology, an endeavor few philosophers

would be fully comfortable with aligning themselves with. But there are infinitely many others as well. Alternative non-natural hypotheses can easily be constructed in structural assimilation to realism. On pain of consistency, wouldn't this move then have to cut a break for all partners, however insane, if it is to do the same for one? Consider realism about witchcraft: the hypothesis that there really are witches who acquire that status in virtue of their natural features; that witchcraft does not imbue their members with any supernatural powers; that it simply consists in having a particular status – being a witch – a status certain persons have by reference to some non-natural standards for witchcraft. No empirical discovery could in principle disprove the hypothesis of witchcraft appropriately refined. Yet I presume witchcraft is a partner realism is better off without.

2.4. The Epistemological Argument

Epistemological worries enable the second most successful way of indirectly assaulting realism by undercutting potential motivations for the view. Since epistemological worries constitute a prelude for what I consider the most successful indirect attack – this one stemming from practical concerns – I start with the former. Epistemological as well as practical worries take their aim at the very heart of realism, namely its independence thesis. By disconnecting normative truths from natural truths, realism inevitably also burns all epistemological and practical bridges required to reconnect the normative to our ways of life. The normative cannot be independent and yet epistemically accessible, or independent and yet practically relevant, or so I wish to argue. Start first, then, with the epistemological concern I hope to substantiate: If there

are normative truths of the sort realism espouses, we are in no position to non-accidentally find out what they are. Left thus clueless, then, we cannot respond to them in ways commonly thought significant to our rationality, morality, or responsibility. I readily admit: it is entirely possible that we may correctly believe, even know, what those truths are. But the chances of that are cosmically minute. The epistemic inaccessibility of normative reality would certainly not disprove that reality; that's why the charge, even if successful, is at best indirect and inconclusive. Still, it most likely would create enough of an aura of discomfort to motivate an interest in my ultimate counterproposal.

The charge itself is quite familiar, occasionally put in terms of the absence of an epistemological story to accompany realism. I am not sure this is necessarily the case; realism, in a sense, does have a story, and therein lies the very problem: the story of blind guessing. The epistemology of realism is akin to playing the lottery. We may hit the jackpot, and there is no principled argument to rule out the occurrence of such a lucky coincidence; yet the odds are rather slim. In fact, the problem does not lie in mere absence of a story at all. There were and still are countless phenomena for which we lack a story. The Greeks had no idea of what drives cognition – they apparently considered the brain's purpose to cool the blood – but that didn't entice them to doubt there is such a thing. The problem, then, is not the lack of a story, but the apparent inevitability of a terrible one.

For the sort of epistemological concern I wish to press against realism to succeed, I need to get the focus exactly right. I must block what I deem are avenues of evasion and questionable retreat. In response to epistemological worries, realists have usually

attempted to explain how normative beliefs can be justified. But the focus on justification, I believe, side steps the deeper and potentially irresolvable issue. The problem is accessibility, not justification. Shafer-Landau at first aptly describes the problem in terms of accessibility, only to immediately turn the lens on justification afterwards: “A classic case against moral realism begins from the assumption that any moral truths there are must be accessible to us, and maintains that realists have no plausible account of such accessibility. To be accessible, in the present sense, is to be capable of being known or justifiably believed.” (2003:231) Not so. To appreciate the true force of the worry, we need to redirect our attention away from matters of justification onto our chances of getting it right. Justification is commonly understood in broadly coherentist terms, as when we reach a state of reflective equilibrium or one that exemplifies our best epistemic efforts. And there are no principal obstacles for normative beliefs to acquire justification in that sense. Yet such a concession does not even begin to address the deeper worry; which is not that we may fail to be justified in our normative beliefs, but rather that the chances of even our best epistemic efforts to get it right approximate zero.

Here, then, is the argument. How could we acquire some non-accidentally correct representation of normative truths? There are three broad ways to cover. We may hit upon the normative in a priori fashion, through empirical investigation, or based on our intuitions. Yet none will suffice to explain how the sort of normative reality espoused by realism could become epistemically accessible. First, forget about the a priori. Realism has already granted that normative truths are not of the logical or conceptual sort; neither are they constructed in virtue of our best attempts to organize our pre-theoretical

normative convictions in coherentist fashion; what is normatively true is not so due to some state of reflective equilibrium we can reach under actual or idealized circumstances. Yet the logical, conceptual and coherent are the prime, if not only, candidates for the a priori. If we do not have a priori access to the normative in virtue of some logical, conceptual or coherentist capacities, how else would that work? Assuming, of course, that we are unwilling to reintroduce some primitive non-natural epistemic capacities widely thought discredited: a primitive moral sense or some non-natural moral intuition that has a mysterious direct line to matters normative.

For sure, this entire problematic would not arise if we could suppose with reasonable confidence that (some) normative truths are self-evident – a path indeed taken by several writers. Unfortunately, we cannot do that. Who would be fully comfortable presuming the truth of utilitarianism, deontology, or virtue ethics, or of hedonism, intellectualism, perfectionism, preference-satisfaction, and so on solely the deliverance of some self-evident a priori reflection? Here's a self-evident proposition: 'If snow is white, then snow is white.' Yet normative contentions, such as that pleasure is (or is not) the one and *only* thing worth striving for are nothing like that. It is one thing to advocate on behalf of the truth of a thesis and quite another to declare it self-evident. Philosophers like to emphasize that reasonable people can disagree about normative matters; which would hardly be so if they turned out self-evident. The persistence of reasonable disagreement supplies prima facie compelling evidence against self-evidence, and there is plenty of that in normative discourse. Insisting nonetheless on self-evident proposition would, hence, amount to the flat-out declaration of intellectual incompetency of many hard-thinking people. It would be accompanied by a presumption reaching barely

tolerable levels of dogmatism that, I'm afraid, also comes dangerously close to reminding us of previous practices of declaring the expositions of opposing viewpoints worth nothing but the fire. I, for one, have certainly spent a considerable amount of time in serious reflection, and yet cannot help finding less and less self-evident all matters normative. Teaching students why matters are harder than they initially appear is in fact an important part of the ethics curriculum. Appeal to self-evidence is thus not only entirely unsubstantiated but also intellectually disreputable.

Next, turn to empirical methods. Natural observation cannot non-accidentally deliver correct representations of normative reality either. If normative correctness is indeed what realists suppose it to be we inevitably are in no position to non-accidentally identify what is correct based on empirical information. This is due to the supposed independence of the normative from the natural – realism's core thesis that a thoroughgoing conception of the natural world does not determine the shape of the normative world. Yet if that is so, we cannot determine what form that shape has based on what we empirically know. Consider, as an analogy, that we have two closed off class rooms; each has written a number on its blackboard. Suppose we previously asked two students to write down whatever number came to their mind; being careful, of course, to ensure that there was no correspondence between the two. Given such mutual independence, there simply is no way to determine what the number is in the other room based on what the number is in this room, apart from just blind guessing. And yet that is exactly the situation in which realism finds itself. We are confronted by a myriad of serious contenders of normative theories all of which equally compatible with the natural. Since every well-worked out ethical theory is equally compatible with what we can

observe scientifically, they are in this regard equally likely with respect to our best natural evidence. To put an artificially low number on it, suppose we have a thousand equally coherent theories of value all of which fully compatible with a thoroughgoing conception of the natural world. What, then, is the basis for assigning even a greater probability to one of them rather than another on grounds of natural observation? Being epistemically locked off in the natural world, how are we to determine what number is engraved in some independent normative world? For all we know, the relevant value-switch could have fallen on each of those thousand theories, and the independence of that result from the natural preempts any natural access to that result.

Turn, finally, to intuitions. It was not my intention to mention intuitions as a third and necessarily distinct epistemic category from the previous two – the a priori and the empirical. Depending on one's account of what intuitions are, they incorporate different mixes of the two. But intuitions now serve as a basis for normative argument so prominent that the extra treatment is warranted; even though what follows will mostly be an application of the previous remarks to the particular case of moral cognition, sentiment and public opinion – to slightly broaden the traditional scope of moral and normative intuitions. Our moral and normative sentiments and intuitions are most sensibly seen as part of the natural order and amenable to natural investigation. There is an increasing number of empirical data shedding light on these phenomena. It is hard to predict what future empirical investigation will reveal; however, in light of what we already know I find it hard to share the confidence of many philosophers that the empirical data will bear out a conception where our moral sense and intuition happens to reliably indicate a causally independent non-natural normative order. While empirical evidence cannot

directly refute such realities – this was the conclusion from above – it can put substantial pressure on the idea that our moral sensibilities could inform us about them by recovering a genesis of those sensibilities that is ill-suited for the task. If those sensibilities turn out to be the product of factors that are not causally controlled by any alleged non-natural moral realities – as must be granted given the way non-natural realism is set up – it would be a cosmic coincidence if the relevant sensibilities were to correctly represent the nature of those realities. In light of the independence of the normative from the natural, the reliability of moral intuitions could only be secured due to some amazing preestablished harmony between the development of our moral sensibilities and normative reality.

Yet what we learn about the genes of our moral sense and intuition in detail does not inspire confidence in such preestablished harmony. That they are shaped by the contingencies of our biological and cultural makeup becomes increasingly hard to deny. It is particularly interesting to note how certain emotional features that are idiosyncratic to our species and culture drive our moral sense (Cf. Jesse Prinz, *The Emotional Construction of Morals*). Requirements that have high emotional resonance – by eliciting disgust reactions, for instance – are more likely to culturally spread and persevere than those with lower resonance (Cf. Shaun Nichols, *Sentimental Rules*). How funny it would be if the contingent agglomerate of causal factors that explains the importance we place on the symbolic, for instance – our tendency to accord more weight and attention to the famous kidnapped baby than to the merely starving baby, or saving private Ryan in the course of which many other soldiers die – just happens to coincide with what is true in a causally closed off normative domain. Or, to take another example, that our predilection towards the near and dear, nicely explained in evolutionary terms by Hamilton's famous

kinship formula, just happens to match what we are normatively required in terms of an independent realm of “special” obligations. Recall the two class-room example from above. The reliability of my intuitive call of what number is written on the other room’s board would be akin to that of rolling a dice. If moreover my intuition was shaped by my idiosyncratic fondness of prime numbers, suspicion would be even more in order. The situation is similar for non-natural moral realism. How to combine a belief in non-natural moral realities with one in the reliability of our moral sensibilities is unclear to me.

With no a priori, empirical or intuitive access to normative reality, we simply must lack such access altogether. Notice, though, that these epistemological worries do not merely reiterate a skeptical point. To treat them as such would be misleading and understating the problem. This can be easily seen from the contrast of normative with traditional external-world skepticism. There we learn that we cannot systematically rule out that all our perceptions of the natural world may be deceptive. The traditional skeptic reminds us of our perceptual limitations – that they are not self-certifying – and that we are consequently anything but epistemically perfect. This we can learn to live with, because what the traditional skeptic cannot take away is our appreciation of how we could in principle form more or less reliable representations of the external world. We have an excellent account, well supported and worked out, of how we may indeed have access to the external world. If there is nothing funny about our situation in the world, we know how we can come to be informed about it. But the epistemic difficulties realism faces are of an entirely different sort. Here we are not merely embarrassed by being unable to come up with a systematic way of proving and substantiating what we already firmly believe: that there is an external world rather than an evil demon deceiving us. It is

not that we merely cannot rule out with certainty some artificial skeptical scenarios. It is that we have a structural argument that threatens all realistic chances of acquiring non-accidental access to normative reality in the first place. This presents an epistemological problematic of an altogether different magnitude.

Now, if that was not already bad enough, things are worse still. My previous argument focused mainly on our inability to adjudicate between sensible normative competitors if realism is true. Yet we are actually in no position to even rule out utterly insensible ones. Why suppose that only suitably coherent theories are to be considered? Who is to know whether normative reality is coherent? We have absolutely no systematic basis for presuming the ultimately correct normative theories to correspond to what we now consider theoretical virtues. Those virtues have evolved through our theoretical encounters with the natural world. Systematicity is a virtue in a causally closed world, simplicity has proven to help predictability, etc. Normative reality may resemble natural reality in that respect, but again, it may not. Perhaps our true ethical requirements are as abstruse as that on Monday's we must eat peanuts and on Tuesday we must climb the tallest buildings. What motivates our confidence that the correct theories would even make sense to us, that we could process them given our contingent provision of computational power, or that we could even fancy them? Moreover, why suppose they will respect our contingent sensibilities? Why suppose what is valuable must be found agreeable by some particular species with its contingent cognitive and conative design features? Ethicists usually look for guidance to a shared and uncontroversial set of principles, such as that every person has equal moral worth, and then proceed on that basis to make their case for more controversial principles. Yet the truth of the former

principles is no more certain than the truth of the latter. Realism is plagued by an epistemology unable to rule out the absurd, to determine as less likely the most outlandish proposal than what has been developed by entire hard-working philosophical communities. If my argument is correct, the epistemology of realism is a disaster of unanticipated magnitude. It threatens the very viability of any serious truth-seeking ethical theorizing.

Assuming, of course, that my argument is correct, and there is no shortage of those who would challenge its soundness. Here I have space to consider only one, namely the concern that the argument over-generalizes, that if correct, it would spell disaster not just for normative epistemology, but philosophical epistemology quite generally. The reply is in a sense an adoption of the partners-in-crime move sketched above; tying up the success of normative epistemology with that of other philosophical enterprises. Shafer-Landau again best articulates the argument. I thus wish to quote him at some length, starting first with his nice summary of the problem to be followed by his rejoinder.

This first argument claims, against the possibility of moral knowledge, that there is no adequate evidence to support our moral views. All the empirical evidence there is is compatible with a wide variety of ethical diagnoses of it. And our only other sources of evidence – our considered judgments about principles and cases – are too dependent on historical, cultural, and personal contingencies to serve as a secure and reliable basis for justifying our moral beliefs.

Yet this criticism, if successful, is sufficient to eliminate the justification we might have for any of our philosophical beliefs. In philosophy, as in ethics, the best empirical evidence underdetermines the correct theoretical understanding of it. In philosophy, as in ethics, we rely very heavily on intuitions and considered judgments to adjudicate between conflicting claims. Try doing modal metaphysics

or analytic epistemology without the use of such convictions. It just doesn't seem possible.

If that is so, then we have good reason to suppose that the following principle is true:

(P) If there is no adequate evidence to support our moral beliefs, then there is no adequate evidence to support our philosophical beliefs. I think that this principle is true, and that we must reject its consequent, rather than accept the antecedent. For the latter course forces a wholesale philosophical skepticism. Given the truth of (P), the price of moral skepticism is global philosophical skepticism. Yet this is surely too high a price for moral skeptics to pay, since they are affirming the warrant of at least one philosophical claim (namely, moral skepticism).

I have two responses. First, the global philosophical implications Shafer-Landau takes moral skepticism to have need further argument. His move is currently subjected to considerable scrutiny, and many have put pressure on its inescapability (Cf. Matt Bedke forthcoming). Not all philosophical enterprises are necessarily in the same boat. Consider philosophy of mind, for instance, which is certainly empirically informed and constrained. In fact, to the extent to which that sub-discipline oversteps empirical bounds philosophers have indeed questioned its reliability. Similar things could be said for philosophy of language, science, physics and biology, etc. More abstract branches, too, such as logic and philosophy of mathematics can again appeal to formalistic procedures unavailable to ethics. Though they as well may rely on intuitions, the intuition that if both A and B are true then A is true is somewhat less troubling than that one must or must not turn a switch in some particular trolley scenario. And when it comes to modal

philosophy, I do find skepticism highly attractive in its own right. Shafer-Landau's verdict may consequently be premature.

But secondly, and more importantly, Shafer-Landau's move is dialectically awkward. The question at issue is not whether normative skepticism is true, or whether we can non-accidentally represent what is morally or normatively the case. Rather, it is whether this would be *entailed* by realism and what in response to make of realism. The question is whether there is an implication from realism to normative skepticism, and it is unclear how the alleged further implication from normative to global philosophical skepticism bears on the truth of that former logical relationship. Alternatively put, even if realism entailed normative skepticism, and even if normative skepticism entailed global philosophical skepticism, this undesirable result would only materialize if realism is true. And that is precisely the target of our present inquiry, however indirect the approach. We cannot here dialectically ignore realism's many rivals, many of which take full credit for their apparently superior epistemology. Naturalists, constructivist, subjectivists, expressivists all believe they can explain moral knowledge with ease, since they don't assume that ethics or normative discourse is aimed at the discovery of an independent normative reality. Especially the practice-based constructivism I favor handily emerges free of the realist's embarrassments of being accompanied by an epistemology of voodoo. Shafer-Landau's acknowledges the very same point: "Constructivists will tie moral knowledge to knowledge of the deliverances of the relevant agents whose attitudes form the basis of moral truth. Obtaining such knowledge will be more or less difficult depending on the particular characterizations of the agents and attitudes that go into constructing moral truth. Nevertheless, the difficulties can in principle be overcome, so

long as we have hope of gaining access to the relevant attitudes.” (231) Why, then, did this opportunity of liberating us from the threat of global philosophical skepticism go unmentioned in the first of the above passages? Constructivism – by the admission of its opponent – does not share the predicament of realism who in virtue of its particular realistic interpretation of what ethicists are trying to achieve potentially lacks any viable corresponding epistemology. Thus, if anything, Shafer-Landau’s allegation that moral skepticism entails global philosophical skepticism would seem to aggravate the burden on realism; realism would now appear even more costly, unless the implication from realism to moral skepticism can be cut, which would bring us back to exactly where we started.

To sum up. Realism, I argued, implies principled epistemological inaccessibility to all matters normative. The point is not of a skeptical, but structural nature. It is the very independence of the normative from the natural that figures as the cornerstone of realism that inevitably involves such dire epistemological side effects. This does not disprove realism, but does burden it with considerable costs. The normative domain – what is valuable, right, or reason-giving – was originally hoped to assist us in making our choices. The normative was commonly supposed to be essentially action-guiding. Yet if we are in no position to acquire non-accidentally correct representations of what is normatively the case, its practical impact vanishes. It must resemble an invisible map, specifying where to go without disclosing what route to take. If realism is true, it would still be the case that some configurations are objectively better than others, that some options are to be disregarded, or that certain responses are favored. Still, we will be in no position to tell and to adjust our choices accordingly. This has most drastic effects in the

case of our inability to detect what authoritatively favors what, that is in the domain of reasons. Our capacity to respond to reasons, many now believe, lies at the very heart of our rationality, responsibility, and even agency. We can still respond, of course, to that which is reason-giving, even with the correct responses, but only without any sensibility to what in fact are the appropriate responses to take. It will be a crap-shoot. Our decision-making could not reflect or express more than a blind guess on our part that this rather than that way was how we were supposed to respond. An invisible map cannot guide our choices. It consequently is not a particularly attractive map to buy.

2.5. The Practical Argument

This section challenges the practical relevance of normative reality as envisioned by realism. I join a rising chorus of writers who have articulated similar worries, including Bernard Williams, Christine Korsgaard, Alan Gibbard, Simon Blackburn, as well as others. Parallel to the epistemological worries presented above, the practical concern takes aim at the very independence thesis of realism. By dissociating the nature of normative reality from what we are actually concerned about, realism also relinquishes all avenues connecting the normative back to our practical concerns. Yet without a foothold on those concerns normative reality becomes incapable of making genuine contributions to our autonomous choices of how to lead our lives. The source of its resonance for our actual choices becomes unclear without a bond to our concerns. Independence purges normative reality of the one job it was supposed to take care of: to

guide us in our choices. Thus unemployed it must sink to the bottom of irrelevance; or so I shall argue.

This would be a most surprising result. Realism, after all, is the view that there are true and authoritative answers to questions of what to do. What distinguishes realism from anti-realism is exactly that it postulates objective resolutions to practical issues. Since normative facts authoritatively settle what is to be done, how can there be any remaining issue of what is to be done? In light of this thought various realists have declared all familiar charges from practical significance against realism confused. The challenger appears to be asking what reasons we have to do what we have reasons to do, thereby portraying nothing but unwillingness on his part to appreciate what reasons are in the first place.

Yet no one is guilty of confusion here. The corresponding question will obviously not be one about reasons. The impression that it must be only reflects a common realistic bias. What undergirds the present concern is an alternative picture of the nature of practical questions and how they are to be answered. We have seen how realists read practical questions: as targeting at an external realm of special facts. By doing so, however, realists profoundly misinterpret the practical in terms of the theoretical, and thereby offer the wrong kind of solution to what is at stake. What supposedly explains what constitutes correct solutions to mathematical problems and to practical problems has the same structure in the image of realism: representing things as they are. The alternative picture expressed by the present worry radically departs from this representational model of practical thought. Correctness in representation is a paradigm ill tailored to practical

thought and choice. By living out our agency we do not participate in some knowledge quiz. What we need is a clear exposition of the alternatives in approach to practical significance – a goal that defines much of what follows later. For now, I shall be content if I can begin to substantiate that there is indeed an alternative to realism's representational model, and that despite all the realists' puzzlement this delivers an intelligible way of asking the question: why care about what we are supposed to care about?

What precisely, then, is the nature of the gap I claim opens up between the normative and the practical as a structural consequence of realism? It is common to understand it in motivational terms – when, for instance, we wonder whether the recognition of what one ought to do is by itself sufficient to provide some motive to comply. And how, in case it does not, we want to understand the extra component necessary to bridge normative cognition with motivation and what role it plays in rational deliberation. Now, understanding the practical in motivational terms is correct in one sense and yet highly misleading in another. In light of a long philosophical tradition that has contrasted motivation with reason we must be careful not to underappreciate motivation. Motivation is a highly complex phenomenon. Anyone who harbors suspicions about reason must ultimately turn the focus of attention back to motivational matters, and think that's where the action is, not in the fantasyland of reason. Employing rationality and reflection is just one particularly interesting way of becoming motivated. I thus need to stress most emphatically: The practical problem is motivational only in the sense in which everything related to choice and deciding what to do is motivational. It is motivational in the sense in which Nagel used the term in his *The Possibility of Altruism*,

mistaken perhaps about the contrast of motivation and reason, but not mistaken in his emphasis that all practical problems ultimately come down to what to do.

This is not how motivation is commonly introduced in meta-normative debates. There motivation is frequently regarded the villain for its potentially inimical relationship to reason. The problem set forth is how to bring in line the recalcitrant motivational parts of the soul with their more rational counterparts. Sidwick writes: “Now we cannot help believing what we see to be true, but we can help doing what we know to be wrong or unwise: thus we are forced to notice the existence in us of irrational springs of action, conflicting with our knowledge and preventing its practical realization: and the very imperfectness of the connexion between our practical judgment and our will impels us to seek for more precise knowledge as to the nature of that connexion.” (Methods, chap.1 third paragraph). This is most apparent when we approach the practical problem from a third-personal rather than first-personal vantage point – thereby grossly misrepresenting the entire problematic that only reveals its true force from a first-personal perspective. When philosophers worry about motivation, they often first confront us with a particularly mean bunch of folk, and then invite us join the exploration of new motivational avenues of getting them to be nicer. If we just could get all the sensible knaves, fools, Lydian Shepherds, Gorgias and Thrasymachus, as well as others to be more respectful of our common ethical codes; how wonderful it would be if we could somehow rationally argue each and everyone to behave. The practical problem thus understood turns into a battle against diverging motivational temperaments of especially obnoxious people. It presents and extension of some imaginary war of us against them now fought out on philosophical territory. All this is rather silly, and does certainly not

represent the worry I have in mind. We thus better put to a side the third-personal problematic and its accompanying motivational caricature once and forever.

Turn, then, to the first-personal deliberative perspective each of us occupies. It is here where the practical problem becomes a real issue – *the* issue. Let me first present the problematic in its grandest terms and then fine-tune it step by step. We all have to figure out what we want from life, what we want all this to be about. We only have one life to lead, our own, and – as far as we are able to keep in check life's many vicissitudes – it is ultimately up to our own choices. *This is what I want*, we say: *being a philosopher, a family-man, a traveler, or what not. This is what makes life worth living for me, which explains why I stick around.* From the perspective of realism, however, the question of how to lead one's own life receives its true answers not in your own concerns but in an altogether different source: in an independent reality that potentially transcends radically what we subjectively deem important. They come from the outside, as it were, and fully coming to grips with what that amounts to is what the practical problematic is all about.

Now, I am fully aware that the question why care about what we are supposed to care about must initially appear puzzling and decidedly unimpressive. This is easily explained. The radical nature inherent in realism of the externality of normative reality together with its practical consequences is easy to overlook. We are accustomed to think that what we are supposed to care about cannot stray thus far away from what we actually do care about. For anti-realists especially, this is an entirely predictable attitude. Normative reality, we believe, is an illusion generated by our disposition to objectify our present concerns. It consequently comes as no big surprise that what we are concerned

about and what we believe we must be concerned about neatly match up, since the latter is a projection of the former. Had triangles been able to form thoughts about reasons – paraphrasing Spinoza – they probably would have construed their nature as being concerned with perfect angles and straight lines. The case in point is marvelously well narrated by a chicken-and-egg story dear to evolutionary psychologists: For chickens, eggs must have magical powers; in particular, they must appear equipped with the peculiar property of to-be-sit-uponness. Chickens must find it unimaginable how one could fail to feel that call. How better to make sense thereof if not by supposing norms authoritative that relate eggs to the one and only appropriate response, namely to hatch them? For chickens, the question why care about (hatching eggs) what they are supposed to care about (hatching eggs) is a real no brainer.

Yet the supposition that our actual concerns must neatly correspond to the concerns mandated by an independent normative reality turns out surprisingly naïve – if normative reality is indeed what realism takes it to be. Especially in light of the above epistemological worries we have no systematic grounds for supposing as more likely than not that normative reality is sensitive to our concerns in the first place; that it is hospitable rather than hostile to what we find agreeable, or, what appears an even more intriguing possibility, that the normative is just plain indifferent to our concerns – as the natural world seems to be. Such a possibility is indeed reflected in various conceptions of value, mostly of an aesthetic nature, such as Nozick's and Leibniz's, where what matters is organic unity, or unity in multiplicity, leaving little space for our all-to-human anxieties. It is unclear, then, why to believe in any agreement between our own parochial human perspective and concerns on the one side and normative reality on the other side.

Belief in harmony, here as elsewhere, amounts to little but an entirely unsubstantiated hope. For all we know, life could objectively be all about jumping up and down as many times as one can, or some other outlandish thing.

In addition, realism's epistemological calamities also undermine the only sensible way I am aware of that would explain why our actual and required concerns systematically converge: namely that we adjust our actual concerns in light of what we recognize are the correct concerns to have according to some authoritative normative standards. However, in the absence of an account for how those standards could reveal themselves to our very natural minds, we simply cannot rule out that the imperatives they issue would have to be judged utterly crazy by the light of our present concerns. Perhaps what we must do is to be as miserable as we can be; perhaps reverse utilitarianism is true, that morally we must maximize pain and suffering. Just think about the odds here. Normative truths are supposedly eternal ones. Think of them as the goal. Now add to this a contingent evolutionary process starting from a contingent set of circumstances, resulting, at one stage, in a complex and intelligent design called *Homo sapiens*. Billions over billions of contingent happenings, neither related to nor influenced by normative reality, added up to the presence of some funny thinking creatures. How probable is it that what they happen to like exactly hits that normative goal? The odds here resemble two people executing independently devised travel plans, one starting from Houston and one from Johannesburg, meeting up in a café of some town in Indonesia – very low indeed. With the real possibility of an unimaginably drastic clash between actual and required concerns, we potentially face a fundamental choice in its starkest terms: to side with our own concerns or to side with those required by an external reality.

Now, what does all this mean if viewed from a first-personal practical vantage point? Here you are, taking great pleasure in spending time thinking about interesting issues and sharing experiences with family and friends and exploring other regions of the world, and yet: for all you know, what makes life objectively valuable might have nothing to do with that at all. It is hard to know how to speculate about normative probabilities, but we have gone through various ways of entertaining the thought that the odds are rather unfavorable that it just so happens that your concerns correspond to your normative reasons. Let's thus consider both cases where they either do or do not so correspond, starting with the latter. Suppose, then, that the good life, as specified by the relevant independently authoritative standards, consists solely in perfecting athletic fitness. I introduced this example in the first chapter, and now shall stick to it. It is not entirely unrelated to prominent Aristotelian ideas that each creature has an inherent purpose or telos, that this telos has a semi semi-biological form as it springs from the nature of things, and that being a good exemplar of its kind consists in perfecting ones nature-given telos. And it is conceivable that athletic fitness could fit that bill. The example is also suitable crazy yet without completely being over the top. And it can be extended: suppose further that normative reasons aim at opportunities for perfecting athletic fitness, and that the theory of athletic consequentialism represents our de facto moral duties and/or underwrites the valid scalar evaluations: the obligation to maximize agglomerate athletic fitness or the comparative superiority of options in terms of their greater realization of athletic fitness.

Now postulate you know all this with certainty. The Holy Grail told you. This forces a choice upon you: to continue to indulge your philosophical, familiar and

exploratory predilections, or to correct your concerns, as far as you can, and strive for bodily perfection. It is no secret where my sympathies lie. And I readily admit: My decision to stay cold in the face of objective duties to radically reform my way of life so as to become a second Arnold Schwarzenegger, the ultimate saint, is clearly and objectively incorrect. It exemplifies a profound deficiency on my part. It represents an unwillingness to act as I should. It renders me a bad and potentially blameworthy person. Yet why would I care about any of these verdicts, all of them coming from the outside, after I so boldly refused to heed the initial requirements? All these additional verdicts do is to restate that which I already decided not to empower with practical significance: the dictates of an external provision of authoritative standards telling me to work out.

Now, everyone concurs we must disregard various ways of how I could come to be recaptured by my objective requirements. Society may get angry at me for my nonchalant noncompliance; I may end up in athletic hell (though perhaps I'd prefer that to athletic heaven). Athletic priests may find ways of guilt tripping me. The point is: this is not how the normative is to engage the practical. I objectively have to become a better athlete, not because of something else I do care about, but simply because that's what I have to do according to authoritative standards. To explicitly eliminate potential sources of distraction, let us grant, then, that no instrumentally undesired effects will fall upon me if I decide to disregard my reasons. So I wonder, and wonder how anyone could fail to wonder, why to follow their prescriptions rather than my own?

I have no clue as how to answer that question. Realism, it seems to me, must simply say that's what you objectively have to do, period. This I already took full notice

of, however, and then continued to ask: why care about what I objectively have to care about? What does that have to do with me and my life? It is not that I will pay a considerable price if I don't, that I will become unable to realize my long standing plans and projects. In fact, as the scenario was set up, quite the opposite is the case. If I was codeliberating with some advisor of mine who took my real concerns at heart, I could see how he could help me in my practical questions; he could make me appreciate concern-related aspects of the options I face that I did not previously notice. Yet in the absence of such concern-related avenues, realism's above insistence strikes me as amounting to little but a heavy and unexplained dose of bullshit, to use Williams' technical term. On realism's own account, there is a further, at the very least logical, step to be taken from the recognition of what one is supposed to do to actually doing it – otherwise all the problems concerning motivation could not even arise – and if one wonders why to take this further step, realism turns utterly silent. Even in the face of objective values, it is still you who has to choose. One may be told all day long what one must do; but it is ultimately your life, and your choice, as how to lead it. The practical perspective is radically yours and simply is not identical with that of any alleged independent normative reality. Realism's incapability to even appreciate that there is a further practical issue is symptomatic of the fact that the answer it offered was of the wrong kind to begin with.

Now suppose that my reasons and concerns do coincide. Authoritative standards call upon me to study philosophy, spend time with family and friends, and to explore the globe. And, since we are still in realism's territory, those reasons are not mere reflections of my concerns, but have a standing logically, conceptually, and nomologically independent of them. Reasons and concerns are distinct entities yet happen to align. Can

the normative thereby come to meaningfully engage the practical? I doubt it. The answer, I think, is brought out by a hypothetical: change your reasons and test whether your concerns follow, thereby turning this scenario into one of the previous kind. And if your concerns do not readjust, as I would reiterate the point from above, then even in the case of perfect reasons-concerns matchup, the most plausible explanation of your ultimate practical choices would not mention your reasons. They would still drop out of the practical picture.

Now, there is a profound worry about this argument, powerfully put forth by Professor Norcross on the occasion of a review of the previous chapter. Here it is. That I have the aforementioned reasons may be metaphysically necessary, and the envisioned hypothetical hence impossible. We cannot test whether my concerns would differ under the alternative reasons-scenario, because there simply is no such alternative reasons-scenario. This is an argument with considerable potency, and I am not sure I have an effective reply. Still, I shall try. First, the question of whether our reasons are, or must be according to realism, metaphysically necessary is unclear. This question raises certain meta-modal issues about which I prefer to stay agnostic. I only want to notice, though, that the potential metaphysical necessity of the normative, as far as we know, could turn out either side. I don't know what we would want to say if, for instance, it turned out that our only reasons are to jump up and down and that this is so necessarily. The metaphysical necessity of the normative appears to raise the stakes of the game. Consequentialism, for instance, would now not only be true or false, but be necessarily true or false. I hope I have occasion to ask Professor Norcross whether he is especially troubled about this latter option. Granted, all this is shadowboxing, and so let's just

postulate for the sake of the argument that first our reasons do align with our concerns in the above manner and secondly that they hold metaphysically necessarily.

In response I shall make two suggestions. A first one introduces various partners in crime – not without noticing the irony of my own usage of the realists’ favorite argument. Whatever its worth, then, let me mention that there seem to be many interesting and non-trivial thought experiments that involve metaphysically impossible scenarios. Some of them are doing considerable argumentative work: if water had been different in chemical composition, we would have discovered that as well. Or, dear to all atheists, if there was a god, the world would be better and involve less suffering, or there would be some evidence of him/her/it, etc. Or, had I been born in prehistoric times I wouldn’t have acquired a philosophy degree from Rice University. All these conditionals seem non-trivially true. We may thus still learn something from considering that which is impossible.

This brings us to my second suggestion. It may suffice that I have epistemological or logical modalities to work with. If the normative is not logically fixed – if alternatives are at least logically possible – I may be able to direct my questions at that kind of possibilities. Moreover, since what I am testing is not normative reality directly, but rather indirectly our responses and attitudes towards it, even epistemological modalities may suffice. The questions I raise all have the structure of what we would do or feel if such and such turned out the case. As long as it remains epistemologically open whether consequentialism or deontology is true, for example, we can meaningfully speculate about our reactions if we found out. It seems correct to say, for instance, that if

consequentialism turned out true, Professor Norcross would acquire more hedons than if it turned out false. Before water's chemical structure was decisively determined, the H₃O camp might have reasoned correctly that in the eventual case of falsification their funding would be cut off. In any case, since our attitudes exhibit a well documented fineness of grain, they may also be capable of sensitivities not limited to what can and cannot be metaphysically the case. I'd be very happy if I could live a thousand years, or fly, or travel back to Kant and beg him to write better. It may simply be a fact about our attitudes that they exemplify sensitivities that can only be revealed by imagining how they would behave even in metaphysical impossible scenarios. Since the practical argument is all about what to make of normative reality, our attitudes and responses, not of normative reality directly, I hope I can draw on a long Fregean tradition allowing for exploring options in thought not corresponding to options *de re*.

The normative, I suggest by way of concluding, leaves practically open what to do, and thereby becomes deliberatively dispensable. Perhaps this only prompts the suspicion that I do not fully understand authority and reason, despite all my functional and structural analyses presented before. I admit guilty as charged, but only under the proviso that there is nothing to be understood in the first place. Conceptually, however, I must insist that I am as clear on *reasons* as realists are – after all, we both converge in our conception. If my argument was contributing to the dissolution of our grasp of reasons and resulted in a profound puzzlement I would consider it a huge and unlikely success. The dialectical situation, again, resembles what hard-determinists conclude in the context of the free-will debate: they believe once you really understand what the notion of freedom of the will involves, you will see there cannot be such a thing – a verdict they

reach in their opinion not as a result of lack of conceptual clarity but for the presence of it. Some authors from the camp of realism, Parfit especially, love to level the charge of confusion against philosophers who have articulated the practical puzzlement about realism; at what point, however, did I slip and committed conceptual blunders? I was explicit that the practical argument does not directly challenge the existence of normative reality. On the contrary, I'd say it leaves intact whatever probabilities there were for such a thing. The argument as I introduced it only attempts to moderate our contingent interest in such a reality. I fully admitted we cannot ask, without confusion, what possible reasons we have to care about what we are supposed to care about, since it was stipulated that there are overwhelming reasons to do precisely that. What we can ask, however, and without the slightest hint of confusion, is this: after clear and careful reflection, after vividly reminding ourselves of everything involved, do we want to choose to act in accord to objective demands, or do we rather want to choose to act in accord to what we actually care about? This is the practical question. Adequately formulating it without lapsing back into representational vocabulary requires quite some effort. The practical question as I envision it is not what I should do or have reasons to do. That issue was settled, and on that basis I went on to ask whether to do what we have reasons to do. The answer to that question, I suggest, simply is not a further proposition but a choice. It essentially involves parking my body in the armchair, with an engaging book and a fine cup of espresso in hand, or to move it into the gym, sweating profusely while being caught in the treadmill. The answer to a practical question is something practical, not theoretical, such as an action, a commitment, a plan, a way of life; it is not a thought that takes the world to be one way rather than another. It does not involve some further belief

regarding the alethic status of a normative proposition. Realists may insist that the nature of the notion of reason is such that once you have recognized a reason you must have closed off all further practical questions. This, however, is simply to reiterate the magic without explaining the trick.

Many realists have distanced themselves from normative naturalism precisely because they find it unintelligible how plain natural facts alone could have the kind of significance we are after. We may fully recognize what is the case yet still wonder what to do in the face of it. The worry I have pressed against realism is structurally similar. The mere provision of additional facts, now of a normative sort, cannot achieve what previously plain natural facts couldn't achieve either: determining what to do in light of what is the case. What else can, then? In one sense, I want to say: nothing, that's the point! Yet in another sense, the answer is surprisingly simple: You can! That choice is possible and how is exemplified by or actual daily choices. Realism is a distraction precisely because it turns into a mystery that which is as clear as water.

With respect to all my funny examples, it could be pointed out that I didn't really believe in reasons to e.g. solely strive for athletic perfection in the first place, and that's why I was able to pump the desired intuitions. This is true. But notice, for the record, that to the extent to which I believe in reasons they are nothing but reflections of our considered concerns. I side with the chickens, though without falling for the temptation of objectification. I also suspect that many people, to the extent to which they find it awkward to entertain the thought that reasons and concerns may pull in different directions, are in fact already attracted to an anti-realist and concern-based portrayal of

reasons. They may think that reasons and concerns necessarily align, but only because the former are some function or other of the latter. To this we will turn when I show how to build reasons, if only of a distinctively non-normative but practical kind. Without realism's independence thesis, there cannot be a gap looming between the two, because, after all, they aren't really *two* but just *one*. With the independence thesis, however, we have all the conceptual resources needed to entertain the possibility that concerns and reasons drastically diverge. And with that, the relevant questions can be asked and the desired intuitions be pumped.

2.6. The Transcendental Argument

What is the transcendental argument against robust meta-normative anti-realism? It is easily stated: If there are no reasons in the fundamental normative sense, we cannot have such reasons to believe in their absence either. If no norm has authority, then neither does the norm to believe what is supported by the best available evidence. This would be so also in the case where that evidence supported anti-realism. How, then, are anti-realists able to make their case, if they cannot rely on any norms, even of the most basic epistemic sort? How, to put the point more generally, is anyone able to argue for anything, if global meta-normative anti-realism was to prevail? A view that appears to entail such dire consequences is better well supported, and exactly therein lies the crux. Global meta-normative anti-realism appears to be cutting off the very justificatory branch it sits upon. Anti-realism seeks to engage a dialectical enterprise while denying its currency; at the same time trying to play a game and yet advancing a move that violates

its constitutive ground rules. No wonder many philosophers have been quick to dismiss the position out of hand!

Let me quote a few recent representative statements that exactly level this intuitive charge. The authors I have chosen already have some affinities for anti-realist positions, or at least take them seriously enough to honor them with detailed responses. The blow of their verdict must consequently exceed that of those already firmly placed on the side of realism. The three authors cited articulate the position they attack in slightly different terms; still, the common thread is quite recognizable. In *The Normative Web*, Terence Cuneo claims that *normative nihilists* – the slightly charged label he uses for anti-realists – face the following dilemma:

Either epistemic nihilists hold that we have reasons to believe epistemic nihilism or they do not. If epistemic nihilists hold that we do have reasons to believe their position, then their position is self-defeating in the sense that it presupposes the existence of the very sorts of entity that it claims not to exist. ... But there are no rational oughts according to epistemic nihilism; there are no facts that imply that certain propositions are belief-worthy or that failing to believe something on good available evidence renders one (all other things being equal) irrational. If, by contrast, epistemic nihilists hold that we do not have epistemic reasons to believe their position, then their position is polemically toothless in the following sense: No one would make a rational mistake in rejecting it and no one would be epistemically praiseworthy in accepting it.

Cuneo thus concludes that this “undesirable result is sufficiently unattractive that any minimally adequate philosophical position will be at pains to avoid being committed to it.” (2007:117-118)

David Copp writes in *Morality, Normativity and Society* (1995:46-47):

No argument can consistently be viewed as justifying normative skepticism, if the argument is also believed to be sound. Normative skepticism is untenable. It would be incoherent to hold that the belief in normative skepticism is justified on the basis of any argument ... For if the belief that no standard is justified is justified, then, since all justification is relative to some justified standard, it follows that the belief that no belief is justified is justified. Hence, the belief is justified that it is *not* the case that the belief in normative skepticism is justified. ... This is a logically consistent position, but it is hardly coherent. ... Worse, it would be inconsistent for a normative skeptic – one who believes normative skepticism to be true – to hold that this belief is justified, for her skepticism entails that nothing is justified. ... This means as well that no one can consistently hold that the belief in normative skepticism is justified unless he avoids being committed to normative skepticism. ... Paradoxically, if one thinks that an argument proves normative skepticism to be true, he cannot consistently hold that the argument justifies belief in normative skepticism.

Richard Joyce writes in *The Myth of Morality*(2001:49-50):

Can we imagine someone questioning practical rationality: “Yes, I recognize that there is a practical reason for me to ϕ , but what is that to me? – Why should I adopt that set of rules?”? This, it seems to me, is incoherent (perhaps uniquely among these sorts of questions). Even to ask the question “Why should I be interested in practical rationality?” is to ask for a *reason*. Thus even to question practical rationality is to evince allegiance to it. After all, what kind of answer could be provided? If the questioner is already expressing doubts about whether things he acknowledges as “his reasons” should move him, then there would be no point in providing further reasons. Therefore to question practical rationality *is* unintelligible – it is to ask for a reason while implying that no reason will be adequate.

Later on, Joyce summarizes the point succinctly: “practical rationality is not something that we may legitimately question, for to question it *is* to acknowledge it.” (2001:83)

The common denominator in the cited sections is that his position is none the anti-realist can coherently defend philosophically. By attempting to do precisely that the anti-realist is committed to the game he questions. By arguing against the existence of reasons in the fundamental normative sense he in fact evinces allegiance to their very existence.

Cuneo, Copp and Joyce are careful to target their arguments only at the philosophical defensibility of anti-realism and not directly at its truth. Their charge is indirect. Transcendental arguments, as Barry Stroud has convincingly argued (1968), cannot establish the falsehood of a position straightforwardly, but at best only that the position cannot coherently be adopted and defended. All three authors acknowledge anti-realism is not *logically* inconsistent or self-defeating. The statement which denies normative facts does not entail a contradiction as does the statement which predicates its own falsity. Neither does it undermine its own intelligibility and coherence by specifying criteria it itself then proceeds to violate, such as the dictum of verificationism which states that every meaningful sentence is either empirical or analytical, without itself being either. Hence, even a successful transcendental argument cannot rule out its target theory’s correctness. What the transcendental argument seeks to establish has more the form of a conditional: If there are no normative facts, then this truth cannot be supported by normative facts either. More generally, if there are no normative facts, then any philosophical debate that aspired to proceed in normative terms would be unsuccessful.

All reason-statements would be false (Cf. Joyce qualifications, though), including that we have reasons to believe so. This result might be regrettable and awkward. We would be caught up in a futile discourse, trying to get at the bottom of things that do not exist. This is how many would regard theological discourse. It would not, however, prove realism true or anti-realism false.

Insisting upon that point, however, would not be an effective way of countering the transcendental argument. For starters, by now we do not need to be reminded that most meta-normative arguments are highly indirect. This reflects the complicated nature of the present discourse where indirect arguments may be all we have. Hence indirectness alone will not discredit transcendental arguments. Secondly, given the greater confidence we take in the sensibility of the philosophical enterprise than we take in particular meta-normative assumptions, the response will be unmoving. Thirdly, what to make of philosophical arguments depends on the larger picture. An affirmative account that does justice to most of what we believe about reasons potentially reinforces the transcendental argument. Since non-debunking accounts are preferable, the anti-realist better does not rest content with his in principle irrefutability. Luckily, he has a more effective response, and to this I now turn.

2.7. The Refutation of the Transcendental Argument

What is the response? I will start with a concession, show why it does not support what the realist wants, and close with a counterproposal that effectively takes anti-realism off the realist's fire line. My goal is not just to score a logical point. The transcendental

argument fails for more profound reasons, and that failure is instructive for it reveals true alternatives in approach. What I contest is not the validity of the transcendental arguments so much as one of its presuppositions. I admit: Justification in the realist sense is indeed not something the anti-realist can draw on, as he rejects it outright. I grant we have indeed no reasons in the fundamental normative sense to believe in their absence and to accept any norms, including norms underwriting rational argumentation. If philosophical argumentation is to be carried out in terms of normative reasons – if the realist gets his way in setting the success conditions for that enterprise – the anti-realist will indeed stand little chance. From the realist's perspective, my above concession must be read tantamount to rejecting all norms that now lay reason-unsupported – including norms all parties need to rely upon to make their case. For he introduced authority as that feature which settles what norms to accept and thus he must consider the non-existence of authority to reveal all norms unacceptable. This reasoning is impeccable, but not inevitable. What it presupposes is an intimate connection between the acceptability of norms and their authority. And this presupposition is exactly what I wish to challenge in my defense of anti-realism.

Rejecting a norm's authority, in short, does not amount to rejecting the norm itself! This sounds paradoxical only under the realist's presupposition just unmasked. The assumption, namely, that the practical issue of what norms to accept must be settled in terms of what meta-normative status norms possess. Disagreement over this assumption leads the realist and anti-realist to see the question of what norm to accept very differently. For the realist, the issue is primarily a theoretical one: finding out what norms bear a certain meta-normative feature. Once that feature is identified in certain norms,

those norms are to be accepted. For the anti-realist, this approach is misguided. He does not believe that the question of what norm to accept is answered for us in virtue of some non-natural property norms possess. He does not believe acceptable norms come attached with some metaphysical seal of quality that would settle the practical issue of whether to accept them. Whether to accept a norm is one thing and what meta-normative properties norms possess quite another. Only the realist, but not the anti-realist, finds the second question to necessarily bear on the first. For sure, anti-realists admit that such a decision has to be made partly in terms of what characteristics norms have; whether or not they serve our purposes, for instance. He at some point owes an alternative story for acceptance. For now, we need to notice that his rejection is twofold. The anti-realist rejects the feature of authority warranting acceptance. He also denies authority to carry the practical significance that realists assign to it. For the anti-realist, the non-existence of authority is in fact no big deal.

To see the difference in approach, consider this thought-experiment. It nicely illustrates the structure of the disagreement as well as the anti-realists' response. I hope the reader will not take offense in its details. Suppose realists thought the acceptability of norms consisted in their approval by the gods, and anti-realists were to deny gods. Structurally speaking, normative authority and divine approval are not entirely dissimilar features as they both purport to provide ultimate backup for norms that with unquestionable authority licenses their dictates. Neither gods nor authority is supposed to be constructed or projected. Now, based on the presumed link between divine approval and acceptance, realists must interpret the realist's denial of gods as a denial of norms. But not so must anti-realists. For they reject precisely that link, and consider the

questions of what norms are approved by the gods (none) and what norms to accept (some) distinct. Anti-realists, in this hypothetical scenario as well as in the real world, have no quarrels with norms, but only question whether they possess certain meta-normative properties. Instead, they picture norms more as tools or strategies for the attainment of our goals, in consequence reversing the master-slave relationship inherently suggested by realism. Norms do not subject us, in virtue of some feature of authority, but are servants for our purposes. We use them, for instance when we are conducting theoretical inquiries or when we are coordinating joint endeavors. We employ norms based on what purposes we have and on how effectively they advance them. If what we seek is truth, we best rely on norms that have proven successful before.

Let me fill in some further details for how I envision theoretical inquiry in anti-realist fashion. Call the theoretical and philosophical enterprise of figuring out what is true *dialectical engagement*, and call the norms guiding the successful execution of dialectical engagement *basic norms for rational argument*, or *basic norms*. All this is admittedly abstract and idealized. Specifying with greater precision what basic norms are would quickly lead into controversies I wish to avoid. My point is a general one and some abstraction shall be unavoidable. I need to be clear, however, on what exactly I assume about basic norms. First, I do not intend to have the locution “norms for *rational argument*” to distinguish basic norms as authoritative, but only to characterize their subject matter as dealing with issues related to how to reason and argue (as opposed to how to garden, say). Basic norms concern cognitive and dialectical landscaping while gardening norms concern backyard landscaping. The term *rational* exhibits an ambiguity familiar from the term normative. It can either be used to merely denote canons of

rationality, prescriptive standards or protocols that could be summarized in some compendiums of *Logic and Reason*. In this case *rationality* denotes just another norm. In contrast, the term can also be used to denote an *authoritative* standard, perhaps even *the* authoritative standard. Many philosophers have in fact interpreted the normative question explicitly in terms of rationality. They understand the question of whether e.g. morality is normative – whether we have reasons to be moral – in terms of whether it is rational to be moral. In doing so they are effectively taking it for granted that rationality has normative authority. In my usage, however, *rationality* extends no further than to denote just another norm. And since I have no worries with norms, I am equally happy to acknowledge a realm of norm-related facts of rationality.

Notice that one Cunoe's above cited contentions is true only under the more ambitious authority-laden reading and false under the less ambitious authority-free reading of rationality. There he says the anti-realist is committed to the view that "there are no facts that imply ... that failing to believe something on good available evidence renders one (all other things being equal) irrational. ... No one would make a rational mistake in rejecting [anti-realism]." As I propose to understand rationality, this would be plainly false. Making a rational mistake is failing to do what basic norms call for. This is not a possibility anti-realists deny or even find problematic. Consider etiquette, the philosopher's favorite example for suspicious norms. Still, even in light of that suspicion, few would think "no one would make a mistake of etiquette in rejecting to wear a certain outfit on a certain occasion." Of course there are mistakes according to etiquette – we can recognize them quite easily. In the case of uncertainty, one may always consult the most recent Emily Post. What philosophers who regard etiquette with suspicion question is not

whether there is a fact of the matter of what conforms to etiquette – there clearly is. Rather, what they contest is whether such mistakes offend anything authoritative (for realists) or whether the thing to do is to stay committed to etiquette (for anti-realists). In similar fashion, certain ways of organizing attitudes are plainly irrational. The inference *I wish that p hence p* is clearly irrational. I reject only that norms of rationality and basic norms are authoritative.

Now, returning to my anti-realist vision of dialectical engagement, I suggested we use basic norms based on how *successfully* they facilitate the discovery of truths. For our purposes, we probably want to adopt a more subjective criterion for success. Norms are successful if for all we know they facilitate the discovery of truths. This again is rather abstract and quick. For now I need to restrict myself to a few comments on the notion that we chose to use norms. I certainly don't believe we consciously chose on any particular occasion to implement basic norms based on their perceived success. I envision the process by which we become attracted to basic norms as rather slow and gradual, where through a constant process of adjustment basic norms are passed on and refined. I would also suspect that our cognitive architecture is to some degree biologically channeled or predisposed to implementing basic norms – despite all popular reports of our allegedly wide-spread irrationality. Whatever the ultimate story, I intended to use the term *successful* as shorthand for whatever characteristics norms have that attract us to them (individually, collectively, and evolutionarily). What I need to emphasize, however, is that I do not claim we have reasons in the fundamental normative sense to comply with *successful* basic norms, or even that we have such reasons in case we are seeking to uncover truths.

My construal of basic norms resembles John Broome's account of normative requirements or norm of rationality. For Broome, norms of rationality are prescribing certain patterns of inference. They require adopting certain cognitive responses given one has adopted others. They require *believing A if one already believes A & B*. The important point for Broome is that normative requirements are wide-scope. They don't call for particular attitudes given ones other attitudes. They call for certain packages of attitudes. Broome does not, however, consider those norms normative in that they do not necessarily issue corresponding reasons or oughts to comply (Broome usage of the attribute 'normative,' I fathom, is meant to denote the directive element in norms of rationality). Broome does not think one has a reason in the fundamental normative sense *to believe A given one already believes A & B*; one is just rationally required to do so. This makes his approach congenial to mine. Norms require (or direct or recommend), and rational norms require to be rational, to conduct reasoning in a certain way, just as gardening norms require to garden in a certain way. Basic norms are norms to organize reasoning about what's the case (that's their subject matter).

Now let me add to my picture the acknowledgment that basic norms not only facilitate, but also enable dialectical engagement. Their relationship is one of constitution and not just one of promotion. In addition to facilitating certain endeavors, norms are commonly known to make some of them possible in the first place. Linguistic exchange can carry on only in virtue of constitutive norms, and so can games and lots of other activities. Some would even include the very possibility of having propositional attitudes or being a cognitive system. Constitutive norms are certainly in play when it comes to the theoretical and philosophical enterprise of figuring out what is true.

However, that a norm is needed for enabling this or that enterprise is not a normative fact, nor does it require one. If in chess one starts moving towers diagonally one has effectively ceased to play chess. If one stops regulating one's assumptions in light of what the evidence supports one has effectively ceased to engage in truth-seeking. The underlying point can be expressed by the following entirely non-normative conditional: one is playing chess only if one abides by chess rules; one is engaging in truth-seeking only if one accords with basic norms for dialectical engagement. There's nothing in this constitutive relationship that would explain or require the presence of any meta-normative non-natural facts. Only the mistaken addicts-understanding of normativity – where something is normative if we can't help using it – could mislead us into thinking that being constitutive for and being authoritative are the same feature when in fact they are very distinct.

As a consequence, there is nothing paradoxical or even incoherent in the notion that the anti-realist may use norms for dialectical engagement in order to advance his position. There's nothing to prevent his participation in the truth-seeking enterprise, just as there is nothing to prevent someone to play chess (and even win!) who happens to question that chess rules are sanctioned by certain non-natural facts (a position, I presume, with few advocates even among chess fans). In case the anti-realist was able to bolster this position successfully – a tall order by anyone's admission – we could say that as far as we can tell his position appears more probably true than not. We would have arrived at this judgment by employing basic norms for truth-seeking to the best of our abilities. We could even agree on calling the position *justified-according-to-norms-guiding-dialectical-engagement*, so long as what we mean by this is no more than that the

faithful employment of norms for dialectical engagement has lead us to this result. Whether in such a case we also would be justified in the more robust sense may or may not be of any concern to us. That it need not be without involving any funny incoherence on our part is what I have argued for. Have I shown that we can have reasons to believe anti-realism? No. I don't believe we have such reasons. Have I shown that the anti-realist can in principle defend his position coherently as the one most likely to be true? I wish. All the transcendental argument shows is that certain endeavors require norms, not that norms have any meta-normative property. Since the anti-realist has no quarrels with norms, the argument seems misdirected.

Perhaps this verdict is premature, though. Perhaps the transcendental argument is best applied to norms directly without the detour over authority. It would then proceed as follows. Rejecting a norm based on arguments requires the employment of basic norms. Such rejection presents no difficulties when it comes to norms other than basic norms, since what we employ and what we reject are distinct entities in such cases. Not so with regards to basic norms. Here we would need the very same norms which we aspire to reject for the argument-based rejection itself! An argument-based rejection of basic norms is possible only in virtue of the employment of basic norms. Here we can recall Joyce' apt way of putting this thought: "practical rationality is not something that we may legitimately question, for to question it *is* to acknowledge it" and "even to question practical rationality is to evince allegiance to it." The transcendental argument, applied to norms directly, would then conclude that we cannot coherently reject all norms based on arguments, for there is (at least) one set of norms – basic norms – that will always be with us. If successful, the argument would show global norm-rejection untenable.

I have no principal worries with the argument as it stands. The argument seems to be doing little more than to spell out a necessary but uncontested condition for the successful execution of a particular enterprise, which happens to be that of dialectical engagement. The condition is the employment of basic norms. Since for the sake of argument I have already granted that this enterprise is enabled only in virtue of constitutive norms, the conclusion of this redirected argument comes as little news. Whatever feature the argument establishes basic norms have, however, it is not the non-natural property of authority outlined above. The argument thus does not address the anti-realist denial. We again must be careful not to confuse the anti-realist with someone who is questioning his allegiance to basic norms. Far from it. The anti-realist holds on to basic norms just as firmly as the realist does. Showing that he can't help it does not contradict his thesis that no norm has authority. He might have suspected anyway that as a matter of design he comes hardwired with certain cognitive dispositions for specific inference-patterns which leave him little choice of cognitive defection. The anti-realist contests only that there really is that feature of authority introduced by realists. Arguing requires breathing too, but that does not reveal the demand *breath!* to have any non-natural meta-normative property. It is just something we must do in order to do something else.

Still, the realist might take the argument one step further. Acceptance and rejection of norms, he might suggest, reveal certain commitments. Thus reformulated, the conclusion of the argument would state that one cannot coherently and rationally abandon one's commitment to basic norms without at the same time retaining the very same commitment to basic norms. Now, setting aside the possibility that one didn't have any commitments either way to begin with, suppose the argument does show that our

commitment to basic norms is indeed non-optional in that it cannot be rationally discontinued. From this the realist might reason further: What is it to be committed to a norm, if not to regard it as authoritative? How to make sense of commitments if not in terms of the belief that certain norms are justified? A commitment to norms without authority seems arbitrary, and worse, if the norm is not even believed authoritative, it seems like a fetish. Yet that is exactly the position the reflective anti-realist wants to take: to sign up to norms he himself does not consider authoritative.

This, I believe, may be the strongest way of putting the transcendental argument. It introduces a new notion – *commitment* – which potentially bridges the gap between accepting a norm and regarding the norm as authoritative. For recall the anti-realist's response to the transcendental argument crucially depended upon emphasizing the gap between the purportedly practical question of what norms to accept and the purportedly theoretical question of what properties norms have. Being committed to a norm is now introduced as the idea of accepting norms in virtue of considering them authoritative.

Not surprisingly, then, the anti-realist will challenge the idea that we can wed acceptance and authority via commitments. And this he can do. Commitments need not be accompanied by any theoretical conviction of the realistic kind in order to qualify as genuine commitments without fetish. What a norm-commitment requires is a stable practical disposition to follow the norm's dictates, a readiness and willingness to be guided by it without regret. Perhaps it also requires a disposition to find certain features salient, or to exclude certain options from consideration. Nor need a commitment be arbitrary if unaccompanied by meta-normative convictions. For sure, we may think of

non-arbitrariness in terms of what is sanctioned by some meta-normative feature of authority. But then the charge of arbitrariness simply begs all questions and retains no independent argumentative force. If non-arbitrary means something else, however, then it is at least possible that norm commitment can be non-arbitrary without being authority-laden. Perhaps this possibility is ruled out more indirectly. But in that case, the notion of arbitrariness needs to be made precise before we can reevaluate the charge. With no clear target in sight, the best I can do is to shoot in the dark. There are many ways for our commitment to basic norms to come out as non-arbitrary: because basic norms do indeed serve truth-seeking best; or because we find the commitment tied up with our identity that involves a substantial amount of curiosity; or because it facilitates other aims of ours; or because it is something we are simply endowed with and that makes us who we are, etc. Suppose we just happen to run by norms for dialectical engagement, perhaps something that is to be explained in terms of biological and cultural co-evolution. This capacity certainly has enabled us to do some amazing things. In order to seriously challenge or in any sense criticize that cognitive capacity of ours as arbitrary in the absence of certain meta-normative facts requires certainly more than the transcendental argument does provide. All this does lead us to larger issues concerning commitment that goes beyond the scope of an already stretched chapter. Notice, though, that in this version the transcendental argument depends upon on a particular way of understanding commitment, one that can legitimately be questioned. Its force now hinges on a controversial and contested understanding of the complex philosophical notion of commitment. This does not show that a further development of the argument will prove

unsuccessful. But it reveals further presuppositions that limit the clean slate and minimalist appeal transcendental arguments are usually supposed to have.

Realism cannot be directly refuted; it is set up that way. This I am fully aware of, and as a consequence I did not aspire to achieve the impossible resulting in only one effect: to provide further ammunition for realists to disprove inconclusive suggested counterarguments – a game I truly wish they would stop playing. What I tried is to make realism unattractive and to reveal it as optional; that there are no principal obstacles for systematically developing and defending a thoroughgoing anti-realist alternative. Whether that alternative is superior to realism cannot yet be determined. For this we first need to see the alternative. If after finishing this long chapter the reader went away with some interest in that alternative, the chapter would have achieved its purpose.

Chapter 3

Against Reductive Normative Realism

3.1. Introduction

This chapter surveys naturalistic or reductive normative realism. The view – conveniently named *naturalism* within the chapter’s confines – holds normativity to be a natural phenomenon. I shall discuss various natural constellations – normative engines – that have been proposed to explain normativity, and show why they must fall short in that aspiration. Desire and its relation to reason will be the most important element in this discussion, since it fuels the most powerful normative engine and provides the blueprint for most others as well. My disagreement with naturalism is vital though subtle: I too shall introduce a naturalistic-constructivist apparatus of my own explaining what reasons and values are all about in due course. What, then, sets apart my approach from naturalism? What does is our diverging assessment of what can and cannot be accounted for in terms of the natural. To motivate the difference, we must recognize first what core

convictions define our conception of the normative and, second, that they cannot be accommodated in the world as we know it. I prefer bold rejection to craven redefinition and consider it a mistake to antecedently constrain our conceptual diagnosis of what reasons are by what we believe exists. For the sake of conceptual honesty and clarity – the core trademarks of philosophy – we better acknowledge what components feature in our conception of normativity and remain frank about whether we can supply a natural foundation for them. What essentially differentiates my own approach from naturalism is that my approach issues a call for reform naturalism finds unnecessary. Naturalism cannot find anything in our conception of the normative to prevent its full naturalization, and it's at this juncture where we part company.

The quibbles I have with naturalism resemble the internal disputes of how best to combat libertarians in the context of the free will-debate. In fact, here as before, the analogy with the free-will debate is helpful, since it provides a strict dialectical analogue to how I envision the present debate. Here and there we find three players naturally paired up into realism and libertarianism, naturalism and compatibilism, and global normative anti-realism and revisionism. Instead of choosing the compatibilist escape-route in response to libertarians – incredibly pretending *free-will* never was concerned with agents originating choice in truly unconstrained fashion – I want to grant libertarians to have identified core components in our conception of free will, but then deny their existence and ultimate practical significance. This leaves the newly rediscovered strategy of revisionism, which on one side acknowledges core libertarian commitments in our conception of free will yet which on the other side substantially rejects their ontology, in this regard siding with compatibilism. The pair of libertarians and realists are

conceptually on target even though there is no target, while the pair of compatibilists and naturalists have successfully identified something that could serve as a target if only it had been our target; leaving only the pair of revisionists and normative anti-realists who realize we must change the target if we wish to hit anything. David Chalmers (2009:22-23) has recently provided a neat structural analysis where this dialectical

...triangle between heavyweight realism, lightweight realism, and anti-realism is found in all sorts of areas of philosophy. ... In each case, the heavyweight realist gives inflationary truth-conditions and holds that they are satisfied, the lightweight realist gives deflationary truth-conditions and holds that they are satisfied, and the anti-realist gives inflationary truth-conditions and holds that they are not satisfied. The first and the second agree on the truth-value of certain sentences, while the first and the third agree on the truth-conditions of these sentences. The second and the third disagree on both of these linguistic matters, but consequently agree on the underlying character of the world: it is such that some parts satisfy the deflationary analysis, but no parts of it satisfy the inflationary analysis. In these cases, one can argue that the difference between lightweight realism and anti-realism is largely semantic.

Naturalism presents a thorny subject to cover. In contrast to non-reductive realism, a position defined mostly in negative terms via its opposition to naturalism and anti-realism, naturalism comprises a variegated set of positive attempts to explain normativity that only share some family resemblance. In my discussion of realism, I afforded the luxury of treating *the natural* as a placeholder without precise specification, since whatever its nature realism held normativity to be independent to *that*. But now details matter. Success and failure in the diverse forms of naturalism will hinge on the particular ways in which they seek to explain normativity. My argument that no form of

naturalism can succeed must raise the concern that it aspires to achieve the impossible. No need to throw the towel just yet, however. If we can identify serious shortcomings in a number of prominent naturalistic accounts that share certain commonalities, at the minimum we acquire some inductive basis for doubting naturalism to succeed in general. In this regard my skepticism that naturalism is able to accommodate normativity is no different from others, e.g. that naturalism is unable to accommodate real magic, even though magic comes in myriad forms. Moreover, my argument and its corresponding diagnosis will be surprisingly transparent with the pleasant result that, if working at all, it can be extended with ease. Ultimately, though, we need to remain realistic about what we can argumentatively achieve in the meta-normative context. While the challenge I am about to present has buttressed my worries about normative naturalism, I admit its verdict is anything but decisive; those more sanguine about naturalism may read what follows as an invitation to point out where I went wrong – e.g. what options I missed – or why further developments may achieve what previous ones did not.

In a nutshell, my worry recasts the familiar one: that there is a conceptual gap no substantive form of naturalism can bridge: from what is to what ought to be, from what is the response we are inclined to take to the one we should take, from what we cherish to what is appropriate to cherish, from what we praise and condemn to what is praiseworthy and worthy of condemnation, from what standards and norms we endorse and accept to which standards and norms we should endorse and accept, and so forth. There's no natural route to get us from here to the Promised Land of shoulds and oughts; no normative engine is capable of delivering the final product of reasons. All we get are communities who find certain ways of living more agreeable than others, who, through a

long process of trial and error, have mostly come to agree to abide by certain norms in order to get along. We may have internalized a considerable number of norms, we may get upset at each other for violating them, we may communicate and debate our shared attitudes, we may decide to apply sanctions to the callous and indifferent, and we may resolve to strive for further reform. But where, then, is that fact, that natural configuration, which explains what ways of live are appropriately chosen as better than others, what norms we should agree to abide by, which norms we must internalize, when we are right and justified in getting upset at others, when it is truly called for to apply sanctions, and what further reform constitutes progress rather than just change?

The problem is not finding standards of correctness; that's easy, so easy, in fact, that we can immediately understand why many want more. Standards of correctness are provided by norms, viz. principles, policies, and prescriptive rules all of which encapsulate a directive element. The issue concerns the status of norms: that there are any privileged norms which possess objective authority. To fully appreciate the issue, we must keep distinct the formal aspect of being directive in character from the substantive one of directing with authority. The first distinguishes norms as norm, but it is the second we need to focus on. Philosophers harboring doubts about the normativity of morality, law, rationality, instrumental reason don't contest the presence of norms in these areas but their authority. Norms of etiquette still serve to illustrate the distinction best. The rule that one must answer in the third person to third person invitations clearly states a demand, but what is its normative force? (Foot 1972:308) Norms are easy to come by; their authority, their status as objective guidelines, however, must appear an altogether more consequential matter. Compare Korsgaard who writes "We live under the pressure

of vast assortments of laws, duties, obligations, expectations, demands, and rules, all telling us what to do. Some of these demands are no doubt illicit ... just social pressure, as we say. ... I call the *normativity* of a law or a demand ... the grounds of its authority ... the way it binds you.” (2009: 2) The fundamental question, then, is whether out of the great plurality and diversity of norms any stand out as objectively authoritative. Naturalists say yes and then aspire to explain why.

3.2. Normative Naturalism

What is natural normative realism? Naturalism, like non-reductive realism, obviously affirms the existence of some normative reality. And again, what renders the view distinctive is of what kind naturalism pronounces that reality to be. The answer, of course, is of the natural kind: since naturalism holds everything to be natural and normativity to be, normativity must be natural. Yet we need to know more precisely what that means. For this we can substantially draw on the previous chapter. There, in reconstructing non-reductive realism, we saw that for naturalism to emerge as a distinct position it does not suffice to acknowledge that all that is normatively relevant must be built up from natural bits and pieces in entirety. What we are after is normative status rather than that which bears normative status: what it is to have this or that status rather than what does have this or that status. Neither does it suffice for naturalism to emerge as a distinct position that normative status supervenes on the natural. Naturalism, in a nutshell, offers a view distinct from non-reductive realism only if it postulates that, in addition to *that whatever is normatively significant* to be natural configurations, *that it is*

normatively significant also figures as a thoroughgoing natural phenomenon. Let's briefly recall why.

Remember, first, the important distinction between normative status and that which has such and such normative status. The previous chapter put great emphasis on the fact that non-naturalists need not deny that everything that has some particular normative status is itself a natural configuration. Naturalists consequently cannot distinguish themselves in virtue of holding solely that the normatively significant fully consists in certain segments of the natural unless they also hold that normative significance itself fully consists in certain segments of the natural. If the good life consists in happiness and health, profitable pursuits and engagement, close relationships and beneficial sociability, insight and understanding, then nothing in that list need give pause to any firm naturalistic metaphysician. The suggestion that there are close relationships, for instance, hardly raises any special ontological concerns. What potentially may give pause to such a metaphysician, though – even with the aforementioned admission – is the idea that there is such a thing as *the good life* – a life particularly worthy of choice, a life we have reasons to strive for, admire, or envy: A proposition entirely distinct from one recognizing that there are lives we do happen to choose, strive for, admire, or envy. In other words, it is consistent to grant that there are e.g. close relationships without also granting that they bear the particular normative status of exemplifying value. The distinction between normative status and that which has it could be taken to separate classical normative ethics from meta-ethics; the former primarily being concerned with what has what normative status, the latter primarily being concerned with what it is to have normative status in the first place.

Now keeping status and bearer distinct is not to deny that they are related. To begin with, if there is no status then nothing can bear that status either. Hence, if normative ethics aims at figuring out what has what status, then success of that enterprise so conceived depends on the viability of a certain meta-ethics. In addition, there are many other ways in which bearer and status could relate. In whatever ways status “attaches” to bearer – a metaphor better used with great caution – this certainly need not be so irrespective of what potential bearers are like. Moreover, insisting on the distinction between status and bearers does not imply that they could ever come apart *de re*. Either because, as in the case of attribute and substance, one never comes without the other; or because, as in the case of essence and origin, they are forged together by metaphysically necessary ties. Pain could be truly bad and never be anything but, and yet, in conception at least, there clearly is a difference between the state and its status. To deny this is to deny meta-ethics and meta-normativity, and to suppose each second-order question to collapse into first-order questions. (Cf. Dworkin 1996)

Consider supervenience next, the thesis that, necessarily, two situations differ in their normative status only if they differ in their natural configuration as well. Though widely considered the hallmark of naturalism, this thesis clearly cannot set apart naturalism from non-naturalism. This is readily apparent once we recall that even dualists such as Descartes were able to firmly and consistently pronounce the supervenience of the mental on the natural without thereby retracting their commitment to non-naturalism. Likewise, there is nothing incoherent in supposing the normative to be firmly non-natural *and* to supervene on the natural. This could be achieved in virtue of God’s maintenance of a pre-established harmony between the normative and the natural; or it could just be

like that, without explanation, but why suppose everything must necessarily receive an explanation? Perhaps our universe contains normative laws as primitive components just it contains basic physical laws as primitive components. Normative laws in that context would be laws that systematically correlate natural configurations with normative status and which would thereby, by fiat, secure the latter to supervene on the former. Leibniz raised the puzzling question of why there is something rather than nothing – why there are laws to govern our universe – which is puzzling precisely because we are at a loss as how to answer it. Rather, it appears there simply is no answer as to why fundamental laws of nature obtain. Likewise, it could simply be the case that there are systematic normative-natural laws where again there simply is no illuminating answer as to why they obtain. That this would spell systematic frustration of our ambition to come to grips with the normative delivers no proof that there is no non-natural normative reality supervening on the natural.

For sure, the thesis that the normative supervenes on the natural could be turned into a distinctive characterization of naturalism, but only if accompanied by other substantial assumptions: namely that the supervenience of the normative on the natural is itself explainable in natural terms. But this would inevitably strengthen the relationship between the normative and the natural and to render it more demanding than mere systematic co-variation; in particular, it would render the normative in some way or other to be a resultant feature of the natural. In fact, normativity as such must emerge as a natural phenomenon. Nothing short of this will suffice to render naturalism a distinctive position.

One way in which the normative could be more strongly related to the natural is via reduction, and at the end of the day it is hard to see how naturalism can consistently refuse to affirm some sort of reductionism. Yet we need to be careful since *reduction* can mean many different things. There is conceptual reduction, explanatory reduction, property reduction – often called type-type reduction, and there is reduction of tokens to tokens. In addition, there is another inherent and telling ambiguity in *reduction*. It can either be read as denoting *explanation* or alternatively as denoting *elimination*. It depends on whether the final product is quite what we initially set out to reduce. The term reductionism is dreaded in our culture precisely because people often are unable to recognize the real McCoy in rubble to which it was reduced. You want candy and instead get celery. You know it's not the same when you see it. Likewise, reductive explanations often change our sense of the relevant phenomena beyond recognition. Naturalists must credibly withstand the suspicion that what they deliver is not normativity but something else. Which, of course, is precisely the kind of suspicion I seek to substantiate in this chapter; that naturalists unconvincingly praise their view as reductive in the first sense while really offer a reduction in the second sense. Normativity cannot be part of the natural world. With no other world to have we must lack normativity too.

There is widespread consensus that *normativity as a natural phenomenon* is not a view committed to the successful reduction of normative concepts to natural concepts; there even is decent consensus that it is not committed to the successful reduction of normative properties to natural properties. Sorting this out could quickly lead us into general metaphysics and semantic theories. I shall not do that here. Rather, let me

consider in brief why the kind of naturalism I am after does not require conceptual or type-type reduction.

Start with concepts first. Here I am basically relying on a long Fregean tradition distinguishing between *Sinn* and *Bedeutung*, or between primary and secondary intentions. Often we are in no position to determine whether or not differing concepts share the same referent on the sole basis of their conceptual content. Moreover, normative concepts are likely semantically embedded in peculiar frameworks containing gross-conceptual connections that cannot be fully captured by natural concepts. Thus, normative concepts may very well pick out the exact same phenomena that natural concepts pick out yet do so “in a special way.” For instance, they may instantaneously render salient the importance certain features have for us that natural concepts may not. Normative concepts as such may contain appraisals of significance not contained in natural concepts, even if ultimately the former and the latter describe the same thing. Consider sweetness: things that are sweet, that is, not necessarily the experience itself. Presumably, being sweet is a plain natural phenomenon, consisting in certain chemical structures interacting with our gustatory sensibilities in a certain way. Now suppose characterization *C* fully describes this phenomenon in purely chemical terms. The term *sweet* and the term *C* may be said to differ conceptually without differing in their denotation. All this could be reflected by the fact that one could have mastery of both concepts yet without knowing that they refer to the very same thing. That being sweet and being *C* are identical would constitute an empirical discovery and not merely a conceptual clarification.

That normativity is a natural phenomenon thus need not be conceptually obvious. Whether some normative situation consists in nothing over and above some natural configuration can be conceptually open even if substantially closed. Under some interpretations of Moore's famous open question argument it must fail because it does not recognize the aforementioned conceptual subtleties – which emphatically is not an admission that the argument is best understood that way or that it necessarily fails. Also, what has been said does not rule out that on occasion we know that concepts of a first sort cannot denote the same phenomena as concepts of a second sort on conceptual grounds alone. Here's a quick proof: Take the concept *non-natural*. This one clearly cannot have the same referent as the concept *natural*. And beyond such obvious cases involving logical operators there are many others as well. A miracle cannot have a natural explanation; God cannot be identical to love; transubstantiation cannot be a chemical process. These observations are entirely vindicated on conceptual grounds alone. One cannot explain a miracle without dissolving it, identifying God with love without becoming an atheist, and provide a chemical analysis of transubstantiation without eliminating it. Likewise, even if we grant that we are in no position to rule out *many* convergences in reference on conceptual grounds this does not mean that we are in no position to rule out *some*. Which is good news for the larger argument to come which substantially relies on some conceptual observations.

Next consider type-type or property reduction. Suppose we think of properties or kinds as suitably stable features that are countenanced by some explanatory theory. A property is what a set of things have in common, and that form of unification is typically achieved and sustained by some explanatory enterprise. Very rarely, however, do

different explanatory enterprises match in what features they pronounce relevant and salient for purposes of unification. Consider money. Being money is a stable feature that figures as a core element in advanced economic practices facilitating exchange. It is hard to specify precisely what money is. But a good guess is that anything can figure as money so long as it performs a certain complex economical function, and is recognized and accepted as such by all relevant participants. Economic theory offers many useful generalizations that proceed in terms of money yet which could not be captured without alluding to money: We know in advance what kinds of pieces this seller will accept in exchange for his furniture and which not, but only if we grasp the concept of money. And then there are reliable higher level economic laws, such that of two kinds of pieces qualifying as money the one considered less valuable will come to dominate as medium of exchange: Bad money drives out good under legal tender laws. This law, formulated by Thomas Gresham in the 16th century and named *Gresham's law* after him, still holds, as we now witness electronic mediums of exchange replacing formerly physical mediums of exchange.

Suppose, then, we agglomerate all things money in a big conceptual pot: gold, silver, certain forms of paper, peculiarly shaped forms of metals and plastic, certain forms of electronic patterns, sea shells, and what not. The important point is that only from the perspective of economic theory will we be able to find those things, and only those things, to share anything significant in common. From the perspectives of relativity theory, chemistry, biology, etc., our class must appear an entirely arbitrary motley bunch. Here you have a piece of paper issued by the federal reserve that is money, just like this coin from Sweden, but there you have another piece of paper, the perfect counterfeit, that

shares all identifiable physical and chemical features with the first but which is not money. How come that this paper and this coin belong to the same category, but not this counterfeit, even though in physical and chemical terms, the two pieces of paper share everything while the federal reserve note and the coin virtually nothing?

This generalizes. In particular, normative kinds could qualify as suitably stable features that figure in certain practices and explanations yet have nothing interesting in common from the perspective of other explanatory enterprises. It is as futile as it is unnecessary an attempt to capture normative kinds in terms of non-normative kinds which are recognized as profitable forms of unifications by other enterprises.

3.3. A Preliminary Case

To warm up, let's look at one prominent naturalistic account of normativity-infused phenomena: reliabilism about epistemic justification. The sketch will be illuminating, I believe.

Take reliabilism in epistemology: the view that belief is justified when reliably produced. This gives us a first prototype of naturalism, and also gives us a first hint as to why it must fail. According to reliabilism, the normatively commendable state that certain beliefs occupy – being justified – just consists in a particular complex natural configuration; it includes historical facts regarding how the beliefs were generated; it includes statistical/probabilistic facts of how frequently the relevant generative procedures result in truths; it perhaps also includes some cognitive sensitivity of the reliability of the relevant belief-forming procedures. The upshot is this: everything

mentioned so far are but plain natural facts. According to this instance of naturalism, being a justified belief is nothing over and above having a particular natural pedigree.

That's all justification *is* says reliabilism – only to inevitably render *justification* normative in name only. What we get is a characterization of a state beliefs can be in, admittedly one we happen to be particularly interested in. Yet if one were to proceed and ask whether we *should* have beliefs that occupy that state – whether it is appropriate to form beliefs fitting that description – reliabilism evidently turns silent. Nothing in the above characterization indicates anything anyone ought to do, has reasons to do, or to become subjected to appropriate criticism if he fails to do, etc. Certainly the mere employment of the term *justification* cannot carry anything with normative significance, even if the term has traditionally been used to with normative pretensions. Call the state resulting from certain process whatever pleases. There are many other states as well one could define for beliefs: being shmustified, for instance, the state of being formed on Mondays, where some beliefs will be shumstified and some will not. For sure, we don't care about shmustified beliefs but do care about justified beliefs, since, among other things, the latter prove particularly useful for building bridges. Similarly, we don't care for bitter fruits but sweet ones, yet *sweetness* no more denotes anything with mysterious normative strings attached than *bitterness* does; it simply identifies something we happen to like. To anticipate the larger argument to come, then, *what we do care about* is distinct from *what we should care about*. Nothing in the reliabilist story appears to provide an explanation of why we normatively must care about justified beliefs.

3.4. Desire, Pure and Simple

Desire fuels the most powerful normative engine. Since it provides the schema and blueprint for many others to come – e.g. the endorsement of norms and the agreement to install and abide by norms – we shall begin with desire. Understand desire to be that pro-attitude which maximizes the plausibility of the relevant response-dependent theory of normativity. Desire, for instance, need not have any phenomenological imprint and usually is not tied up with sexuality. Later I shall argue that the kinds of pro-attitudes that stand the best chance of carrying reasons are concerns embedded within our larger identities and projects. For now, the basic idea is that with the appropriate provision of desire comes reason and value; that practical reasons enter the stage with agents who harbor concerns and adopt ends. The universe is cold and indifferent, but we are not, and in light of our ends and what we care about some practical pathways emerge as more attractive than others. In the final analysis, it is supposed, reasons and values rest in desire, not the other way around. Spinoza wrote “It is clear ... that we do not endeavor, will, seek after or desire because we judge a thing to be good. On the contrary, we judge a thing to be good because we endeavor, will, seek after and desire it.” (Spinoza, Ethics III.9 Scholium) And Simon Blackburn adds “Nature itself may be heartless and free of desire, but among the creatures it has thrown up are some which are not heartless, and not free of desires. We understand our values by understanding ourselves as valuing, and this we can do.” (1998:50). I share a great deal of sympathy with what Spinoza and Blackburn have to say. In fact, reading their statements carefully, I fully agree. Spinoza talks about *judging a thing to be good* and not *a thing*

being good; and Blackburn talks about *understanding our values* and not about just *values*. I too believe desire fuels under perception of and response to things as good, worthwhile, and reasonable. It is what anchors and sustains our practices. If that was all we seek to understand, I am on board.

But it is not. Usually desire is called upon to explain reason and value in their distinctive normative sense; not just our perception and thought about reason and value, but reason and value, period. And here I must depart. I find the desire-based account of normativity a view astonishingly close for being entirely off target. Normativity cannot be explained on the basis of desire. Desire-based accounts deliver something, and we may choose to call it *reason* and *value*. Given their wide employment to denote most everything, we may choose to allow those terms yet another usage. But labeling should not deceive us that the thing named entirely lacks anything distinctively normative. What we need is to look inside the engine and to see what it actually does. We shall see then what it explains is why we choose what we choose and why we prize what we prize. What we shall miss, however, is why we have reasons to choose what we choose and why we have reason to prize what we prize, or why doing anything of that sort can ever be appropriate or inappropriate.

In my criticism I partially converge with non-reductive realists who too doubt that reasons can be explained on the basis of desire. This may seem puzzling, since I have already rejected what they offer, namely to take reason and value as primitive and simply as a given without further explanation. With this option ruled out, then, what else is there to rest reason and value upon if not our concerns and ends? Nothing, I say, and that's

precisely my point. I understand, however, why this move is dialectically unusual. The context of the debate has largely been defined in terms of the contrast between desire-based and more robustly realist accounts of normativity, e.g. value-based accounts. The implicit suggestion, then, was that it must either be the one or the other: Either we accept reasons with no further explanation or we do provide an explanation for them. Since many share a profound skepticism about robust realist accounts of reason and value – to simply take reasons as a given – desire-based accounts have found an important spring of motivation in the perceived absence of viable alternatives. They came to appear attractive for their comparative advantage over realism with its insatiable hunger for mystery. The problem, unfortunately, is that desire-fueled normative engines do not work either, which is a matter entirely distinct from what we happen to think of robust realism. One cannot just rest content with showing stronger proposals deficient; one needs to show that the weaker ones are not deficient as well, and that task often fell off the sideways.

There are desire-based accounts of reasons and there are desire-based accounts of value. Since both proceed in different manner they need be treated separately. I shall mostly focus on reasons and then briefly turn to values. For now, recall what we are after. What we seek is an explanation of how states of affairs can stand in the reasons-relationship to choices agents face in particular circumstances; how certain states of affairs can favor certain responses. The relationship itself needs explaining and not just what happens to stand in that relationship: which is just plain situations, agents, and choices. What favors what hardly gives rise to any profound puzzlement. *That* something favors what it favors, in contrast, is what it is hard to understand yet what desire-based accounts need to explain. Sporadically one finds authors wondering what the fuzz about

reasons is all about. Reasons are just plain state of affairs and situations, they say, and what's problematic with that? Nothing, of course. Yet situations are just situations and are reasons for anything only in virtue of standing in the favoring relationship to whatever they favor. How, then, do desire-based accounts go about explaining reasons and the corresponding favoring-relationship?

Jonathan Dancy proposes an excellent précis of desire-based accounts of reason, where p is some state of affairs, A some agent, ϕ some practical option or choice, and e some end or concern (2000:28):

If its being the case that p is a good reason for A to ϕ , *this is because* there is some e such that A actually desires e and, given that p , ϕ -ing subserves the prospect of e 's being realized (or continues to be realized).

Dancy is no fan of desire-based accounts, but his characterization is remarkably close to Mark Schroeder's, who is a fan. Suppose r stands for some proposition (or state of affairs), x for some agent, a for some action (or option or choice), and p for some desire, end, or concern. Schroeder writes (2007:29):

For all propositions r , agents x , and action a , if r is a reason for x to do a , that is because there is some p such that x has a desire whose object is p , and the truth of r is part of what explains why x 's doing a promotes p .

That Dancy talks about state of affairs (or what is being the case) and Schroeder about propositions need not distract us here. These are variations in detail we can safely ignore for present purposes. The important upshot is that according to both accounts, reasons are complex quadruple relationships between (1) state of affairs, propositions or situations, (2) agents, (3) actions, options, or choices and (4) desired ends or concerns.

Or, to put it differently, for something to be a reason is for it to stand in this complex quadruple relationship. Schroeder provides a neat example (2007:1):

Tonight there is going to be a party, and everyone is invited. There will be good food, drinks, friends, chat, music – and dancing. Ronnie and Bradley ... have been invited to the party. But while Ronnie loves to dance, Bradley can't stand it. ... So while the fact that there will be dancing at the party is a reason for Ronnie to go, it is not a reason for Bradley to go. ... Ronnie's and Bradley's reasons therefore differ. ... Moreover, it's not hard to see why Ronnie's and Bradley's reasons differ ... It is because of what they *like*, *care* about, or *want*.

Ronnie's having a reason thus consists in the quadruple relationship between a situation (the party), an agent (Ronnie), an end (love of dancing), and a action or choice (going to the party) such that the situation causes a particular choice to be the (best) option for an agent to bring about what he cares about.

Now besides Dancy's and Schroeder's version of desire-based accounts of reasons there are others as well. There is in fact little consensus on how precisely desire is supposedly implicated in reason where one can find multiple and conflicting accounts in the literature. Yet in my assessment Dancy's and Schroeder's version fares comparatively well and avoids multiple problems overshadowing many of its competitors. I shall henceforth stick to it. Since the account is able to claim support of prominent proponents and opponents alike, this choice shall not prejudice the discussion to come. Let's start by drawing attention to a few of the account's key features. It is vital to get the picture right. Philosophers tend to be somewhat impatient with desire-based accounts and as a consequence identify flaws that aren't really there, or at any rate needn't be there. When suitably worked out desire-based accounts have in fact a lot going for them. Their strength shall not go unmentioned here. In particular, I shall make five

comments on behalf of the present version: (1) It can adequately capture the deliberative perspective of agents and (2) it can motivate many concerns we have even in the absence of a basis of assessing fundamental ends and (3) it can do justice to the epistemological complexity of practical reasons and (4) it can withstand most attacks by counterexample and (5) it is motivated by a requirement on practical reasons many accept. With a representation of desire-based accounts in their strongest suit we shall finally turn to why they nonetheless remain incapable of explaining what they set out to explain, namely normative authority.

(1) Notice, first, that in the present version it is not desires themselves that *are* reasons. Instead, desires figure as components and constitutive parts of the relevant reasons-relationship, though certainly important components and parts. Reasons surely are complex. Where there are reasons, according to the present account, there are situations and agents and ends and choices appropriately interrelated. If we wish to single out one of the above components as *the reason*, that should be states of affairs and situations. This choice would comport with common usage where we often call the relevant situation itself *the reason*: as when we say the dancing at the party was the reason for Ronnie to go. This way of speaking is perfectly fine as long as we do remember that for reasons to be reasons – for state of affairs to be reasons – they need to stand in that particular favoring relationship. And to favor, according to the account in question, is to stand in a quadruple relationship that only features desires as one important component.

Similarly, we best avoid speaking of desires as “generating” or “providing” reasons. This locution only suggests a rather misleading temporal picture involving two stages, where first there are desires and then thereafter, in some mysterious manner, they “generate” or “provide” reasons. Desire-based accounts essentially are accounts of what reasons and the relevant reasons-relationship are: take them apart, and this is what you find: a quadruple relationship connecting situations, agents, ends and choices, and certainly not a process of dubious creation. Schroeder, in his general analytical approach to understanding reasons, compares his account to the analysis of water as H_2O . There as well it would be rather awkward an expression to say that the chemical complex H_2O *generates* or *provides* water instead of just being water. If again we wish to single out one component as the provider of reasons, it would be as before state of affairs and situations. It is the dance-party which provides the reason for Ronnie to go. (Likewise, it is hoses and faucets which provides water and not H_2O).

These are not merely terminological reminders but have larger ramifications. Reasons are world-wise and practical deliberation consequently is outward-directed rather than inward-directed. Deliberation engages the world and rarely amounts to a mere exercise in navel-gazing. Desire-based accounts do not face principal obstacles in acknowledging these fundamental facts so long as they stay clear on how precisely desire interacts with the other components and, most importantly, the world. What recommends certain choices instead of others first and foremost has to do with how things are. It’s the dancing at the party that makes all the difference for why Ronnie should go. This is not to diminish the important role of desire, for without it no situation by itself does ever recommend anything. Yet as important the role desire plays, the role is better seen as that

of a background enabler, which is that of a spotlight operator rather than that of a frontal stage actor. We look at our options through our desires instead of paying them direct attention. It is in light of one's desires that certain practical options become more attractive than others, where importantly the light springs from desires instead of shining on desires, and where what is illuminated in particularly favorable colors are choices rather than desires.

Only so conceived do desire-based accounts stand any chance of successfully capturing the deliberative perspective of agents. Agents are the unique consumers of reasons. To them reasons must be addressed and they must determine whether to take seriously what is so addressed. That's why it is such an important test case for accounts of reasons that they not misconstrue the perspective of us qua deliberators, the chief arbiters of reasons. Desire-based accounts would fail in this regard if they placed the primary focal point of reasons on desire. Fortunately, they do not. To desire is to take a stake in things, but what we take a stake in are things and not desires. Deliberation very rarely concerns what to desire rather than what to do given how things are. Here's an example. Desiring to have a piece of chocolate-cake is to look at that piece in a certain manner: finding it attractive, taking pleasure in the thought of eating it, and then figuring out how to get it. It is not that one has a neutral perception of the cake, and also notices a desire somewhere to eat it, on the basis of which one then draws the practical inference that one shall eat it. If this was the way desires came upon us, it would leave it rather obscure why we should care about them at all. Suppose I was to discover my desire for chocolate-cake, say by locating it on the computer-screen of some MRI device I am presently plugged into or by some process of psychoanalysis. That result may strike me

as quite interesting; as yet another curiosity about myself. But that fact certainly is not what drives me to get that cake; what does is my fondness of it – rather than my notice of that fondness – with the characteristic accompanying outward way of looking at the cake.

Indeed, to think of desires as the direct object of deliberation immediately leads to a number of awkward results. At one extreme we find desires being presented as if they were a nuisance to be quieted, where one primarily deliberates in order to seek out ways of shutting up the constant grumble. From this perspective it appears as though to desire is to suffer from some mild sort of itch one has not yet figured out how to scratch. This view is bizarre but not uncommon. Jonathan Dancy, for instance, credits Brad Hooker with the suggestion that “on the Humean showing we have as much reason to abandon a desire which there is no reason to have, as to do what will subserve it.” (Dancy 2000:39) Yet only the caricature of the sort outlined above could impose this absurd result on desire-based accounts.

In fact, desire is anything but a self-obsessed creature demanding substantial attention and which above all else aims at its own *satisfaction*. Here, then, we find a first hint as to why the picture likening desires to conditions we need to be cured of is so common: it is concealed in thinking of desires as something to be satisfied, and hence the usual label *desire-satisfaction* is hardly as innocent as often supposed. It potentially steers us in the wrong direction from the very start, advancing desires as something to make go away; as if desires were entries in a list of errands to take care off; as if Ronnie, returning from his dance-party, now had one thing less to worry about. Yet the satisfaction of desires as such is never what we have reasons to do, unless we are very careful here and

understand the notion of satisfaction entirely *de re*; that is to say what we have reasons to do is *what* satisfies desires, *de re*, not to satisfy desires *per se*. Even this may be too much of a concession, since the concept of desire need not even figure in the description of the reason-supported choice at all. Griffin offers the best case in point: "... if a father wants his children to be happy, what he wants, what is valuable to him, is a state of the world, not a state of his mind; merely deluding him into thinking that his children flourish, therefore, does not give him what he values." (1986:13) Again, desire-based accounts are most plausible when desires are merely taken to constitute the explanatory background of reasons and are not considered part of what we have reasons to do. In this regard, the reason-desire relationship is analogous to the money-attitude relationship. Our attitudes towards money explain that it is money, but no piece of money contains any attitudes as proper part. Likewise, desires are taken to explain what we have reasons to do, but reason-supported choice needs not contain desire as proper part. (Cf. Korsgaard 1996; Smith & Pettit 1990)

Perhaps there is another reason why the misleading picture of desires *qua* mild irritations appears so common. And this may have to do with the term *desire* itself. In our culture especially, desire has not been looked upon favorably, in part for its connotation of sexuality and our culture's uneasy relationship with that. In this context being "free of desire" was considered a desirable condition. For these and other reasons philosophers have been looking for other labels, such as the boring technical term *pro-attitude*. I believe *concern-based accounts* would be the best terminological choice, and I shall later say more why. For now, let's just remind ourselves that the sense of desire that is capable of sustaining sensible versions of desire-based accounts of reasons must involve true

concerns for how things go and for how people fare; desiring essentially is a way of getting involved, of becoming less indifferent about the state of the world. To the extent to which felt satisfaction in ends achieved reflects something essential in those ends – which it usually does not – the satisfaction again is world-wise: gratification that one's institution prospers; that one's children flourish; that one's dreams became true. Decidedly not that one finally has put to rest an irritating inner voice calling to serve one's institution, take care of one's children, or to strive towards the realization of one's dreams.

Now it is true that deliberation on occasion does pay direct attention to desires instead of how things appear in light of them. But even then the deliberative result is surprising and quite revealing. A few examples should make this plain. That one desires to get drunk more often than not is a reason to stay away from the bar rather than to gravitate towards it; that one desires to be loved by everyone more often than not is a reason to seek moderation and retreat rather than company; that one desires to take risks more often than not is a reason to stay clear of risks rather than seeking them, and, for instance, not to become a pilot.

When we attend to desires directly, this often is an indication that something is amiss. When, for instance we deal with obsessions and addictions. There is a long tradition in philosophy that presumes the main focus of the deliberative lens to rest on desire; as when we decide what desires to have, to endorse or identify with, etc. The problem that motivates this shift in focus often has to do with attempts of dealing with alienating urges that are characteristic of drug addiction. Ironically, by doing so they tend

to render desiring into a phenomenon too much resembling addiction altogether. For now we need to bear in mind why this purportedly self-obsession of desire is not forced upon us by desire-based accounts. Our desires are part of who we are, and usually figure as a given on the basis of which deliberation can take off. Very rarely is desire itself scrutinized in terms of reasons, and even more rarely can we affect what to desire in a manner comparable to which we can affect what to do in response. I have always been very fond of small company and have despised large company. Perhaps it might have served me better to be more of a party-animal or to at least to be more tolerant of huge crowds it, but the disposition of mine to stay away from large crowds has never been the target of any serious thought. How could I possibly change it, and why would I even want to if I could? It's simply how I am. The only question that usually plagues me is how to get home as quickly as possible.

(2) A very important aspect of the account, then, is precisely that there is no basis for assessing fundamental ends, but only options as facilitating those ends in better and worse ways ones those ends are in place. Still, there is a deeper worry this generates. Even if according to desire-based accounts only means but not ends figure as appropriate targets of assessment, the concern is that means cannot obtain a commendatory status that the ends towards which they are tailored entirely lack. It is frequently objected that without reason to support desire, desire cannot support ways of realizing desire. "If I had no reason to want to catch this train" Derek Parfit writes, "I would have no reason to leave now." (2011 Chapter 1, 29). And Warren Quinn adds "If my basic love of music doesn't give me a reason to listen, then it doesn't ... give me a reason to take the record

down.” (*Putting Rationality in its Place*, quoted from Dancy 2000:32). And Christine Korsgaard must have a similar point in mind when she writes (1997:223; 2008:35):

The instrumental principle, because it tells us only to take the means to our ends, cannot by itself give us a reason to *do* anything. It can operate only in conjunction with some view about how our ends are determined ... It is routinely assumed ... that ... our ends will be determined by what we desire. But if you hold that the instrumental principle is the only principle of practical rationality, you cannot also hold that desiring something is a *reason* for pursuing it. The principle, “take as your end that which you desire,” is neither the instrumental principle itself nor an application of it. If the instrumental principle is the only principle of practical reason, then to say that something is your end is not to say that you have a reason to pursue it, but at most to say that you are going to pursue it. ... If we allow reason a role in determining ends, then the instrumental principle will be formulated this way: “if you have a *reason* to pursue an end then you have a reason to take the means to that end.” But if we do not allow reason a role in determining ends, then the instrumental principle has to go like this: “if you are *going* to pursue an end, then you have a reason to take the means to that end.”

And it is this second formulation Korsgaard is very suspicious of, for it all too conveniently attempts to derive an ought from an is. In sum, whatever normative force desire is capable of passing on must supposedly be borrowed from reason supporting it, and hence the worry is desire cannot be the complete story.

Consider this case in point. Suppose I desire to ϕ , and suppose ϕ -ing avails itself as a direct way of realizing what I desire, namely that I ϕ . The pathways towards realizing desires are usually not that direct and involve more sophisticated navigations

with multiple steps. But sometimes it is that simple and means and ends do coincide: I want to listen to Beethoven, and this is I best achieve by listening to Beethoven. Consider this a limiting case. Now, for the sake of the argument, let us grant I have no reason to desire to ϕ . Yet because ϕ -ing presents itself as the unique opportunity of realizing what I desire, I now have a reason to ϕ , according to the above model of desire-based accounts. Solely in virtue of desiring to ϕ and the fact that I can achieve this directly I have a reason to ϕ . To spice up the case even further, consider intention. Acting involves intentions, and intentions are often understood as states involving desire or resulting from desire. Thus, since doing ϕ involves forming the intention to ϕ , and since this intention includes or results from the desire to ϕ , I automatically acquire a reason to ϕ solely in virtue of intending to ϕ . This sort of bootstrapping appears suspect. Derek Parfit considers one concrete case, the desire to be happy. “The fact that we had this desire could not be truly claimed to give us a reason to have it. Desires cannot be self-supporting. Our wanting happiness as an end could not give us a reason to want happiness as an end.” (2011 Chapter 1, 38)

And yet, it remains unclear just how damaging this problem really is. Recall, first, that desire-based accounts have no aspiration of explaining reasons in a vacuum, but always only relative to particular ends. *Given that this is what you want*, we hear the advisor saying, *here is a good way of getting it*. Never is the corresponding recommendation just to do such-and-such, period, but rather *to do such-and-such given this is what one wants*. It would be mysterious indeed if one could start with some reason-unsupported desire to ϕ and then end up with some reason to ϕ , period, completely

detached from the former desire and standing on its own. But nothing of this sort is suggested. According to desire-based accounts, reasons are always reasons vis-à-vis ends. And bearing this in mind, the supposition that one has reason to ϕ given that one desires to ϕ does not sound as counterintuitive after all: given that I want to listen to Beethoven, I better listen to Beethoven, rather than, say, Britney Spears. For sure, the advice to ϕ if one wants to ϕ does not provide helpful advice, as good advice never is that obvious, but that's an entirely distinct matter. And with regards to Parfit's example, desire indeed cannot support itself. But even when one can ϕ in order to bring about what one desires, ϕ , our present account does not imply that one then has a reason *to desire to* ϕ . That would indeed be an amazing sort of self-certification. All we have, in such a case, is a practical reason *to do* something, namely to ϕ , which is a different thing entirely from reasons for the respective attitude itself.

Desire-based accounts, then, have no aspiration of providing a basis for assessing fundamental ends. They do, however, provide a basis for assessing non-fundamental ends, which is to say most all of the ends we adopt. Desires and concerns permeate agency in the form of integrated organic structures. In most cases our concerns are systematically interconnected, as when we care about something because it possesses features that attract us; as when what we care about is embedded within our larger identities and projects, not necessarily in relation of means and ends, but in one of many other ways. Reading good philosophy is not primarily a means to becoming a philosopher rather than a way of being a philosopher. If one came to regard it a necessary chore to advance this particular career its prospects would be rather dim. And so when we love

reading good philosophy, we do so partly because we enjoy it; because it contributes to our understanding; because it is part of larger research projects; because we need something to argue with. This could obviously go on forever. Indeed, hardly ever are we short of answers why we care about what we care about. Not primarily because we can locate the relevant concerns within some sort of chain with some ultimate regress-stopper. Instead, what we find has more the form of an organic clutter of variously interrelated concerns. The fundamental/non-fundamentals dichotomy is thus best seen as a technical device to make a technical point; not to suggest our concerns need be layered, but to insist that almost never do they just stand in isolation. Sometimes, though, we will say we just don't like it and that's the end of it: perhaps we just don't like feeling pain or getting sick. What is it about pain we don't like? Well, it's just like that, we don't like it. Hence, desire-based accounts have no difficulties acknowledging that our desires are usually motivated by some of the valuable properties of their objects; the value of these properties, of course, must in turn be explained by some other of our responses towards those properties. Saying that *X is desired* may thus not be the end to the story. Thus, even if, in principle, the story needs to end somewhere, and, according to desire-based accounts, has to end with some desire rather than some unexplained value, most desires simply do not stand as isolated givens.

(3) What our reasons are is usually not an epistemologically straightforward matter. It involves lots of intelligence. Desire-based accounts need not implausibly endorse the transparency of all practical reasons. Not merely because what we desire often is anything but translucent to ourselves. Even if our desires were an open book to ourselves, desires still engage a complex world. And what we have reasons to do just as

much depend on the state of the world as on the state of our mind. The complexity of the situations in which agents find themselves, together with the inscrutability of large parts of the future and the exact effects of our choices directly translates into the complexity of practical reasons. And with that comes ample opportunity to discover what our reasons are, to disagree about what they are, and to be mistaken about them. What best sub-serves our ends more often than not is a highly complex matter. How best to negotiate through academia, for instance, so as to realize the end of attaining a doctoral degree is something lots of thought and discussion can be spilled upon.

Notice, however, that the mere fact that we can be mistaken about desire-based reasons does not render them normative. The possibility of mistake about some phenomenon does not render the phenomenon itself normative in character. At best it renders cognition about the phenomenon normative, but that's an entirely different matter. We often are mistaken about what time it is, but time, whatever else it may be, is certainly not a normative phenomenon.

(4) The most common strategy to combat desire-based accounts is by counterexample. This can quickly generate quite extensive discussions. Here I shall stick to a few comments. Overall I remain unimpressed. More often than not the proposed counterexamples fail to fully appreciate the difference between what we have some reason to do and what we have most reason to do; between pro-tanto reasons and all-things-considered reasons. This distinction, in some form or another, is accepted by virtually all parties, and so it is only fair that proponents of desire-based accounts rely on it as well. If someone, to take an example, were to stand on a cliff and harbor the wish of

finding out how it feels to freefall, desire-based accounts may indeed imply some reason to jump. *May*, I say by way of qualification, because for subtle reasons addressed shortly it may also be the case that it does not imply that. For the moment, however, and for the sake of the argument let's just grant that there is some reason to jump. This anyway is a concession entirely safe and sound, given that there must be overwhelming reasons not to jump as well – supposing our agent cares about life and limb at all. His reason to jump need not cause worries concerning his safety as long as it is kept in check by overwhelming reasons to stay put. That we may want to accord him with a reason to jump that only is rendered all but invisible by the opposition – as opposed to according none at all – is supported by this consideration. Cancel the opposition and see what happens. To that end we shall add a bungee-cord. Thereby we may turn into a winning reason what before was squarely on the losing side; and yet the reason, it is plausible to suppose, was there all the same. The bungee-cord hardly added a new reason that was absent before but rather just cancelled out the competition, leaving the field entirely to the jumping reason. This appears the most natural and non ad-hoc diagnosis of what changed the overall normative situation.

And yet, the above concession is not forced upon desire-based accounts, addressing now why our agent need not have a reason to jump after all. We need not presume each and every desire to have the capacity of empowering the relevant reasons-relationships. This is so even if we restrict the focus on desires people actually have, an assumption that has guided the discussion up till now. Desire-based accounts are entitled to add further conditions on which actual desires count. I already have voiced my conviction that only desires that are embedded within our larger projects and identities

have reason-related capacities. Other plausible restrictions could include that the objects of desire must have received at least some consideration; that the desires must not rest upon our attraction to features that are not really there but which we deceptively only take to be there; and so on. None of these qualifications need be ad hoc or threaten our preoccupation with desires people actually have. Hence, more often than not it is dialectically an impotent move to advance just any desire and then to insist that desire-based accounts must acknowledge a reason accompanying it which is of an allegedly counterintuitive sort. It's simply not that simple.

All in all, I share a great deal of sympathy with desire-based accounts. For convenience, we may call them *Schroeder-reasons* or just *S-reasons*. They comport with a substantial proportion of the term's common usage, as is nicely illustrated by Schroeder's example of Ronnie and Bradley. Furthermore, they capture situations of agents in a manner that reflects their own appraisal of these situations; what they find salient and significant about them and what choices will consequently appeal to them. By keeping close to agents' own deliberative vantage points S-reasons afford vast predictive powers. They are tied up with the intentional stance that unlike others is actually able to issue reliable long-term forecasts of what people are going to do. In this regard S-reasons' predictive utility may presently be unsurpassed. This is a pleasant benefit when we occupy a third-person perspective on other people. But desire-based accounts are in fact most attractive when we occupy a first-person perspective on our own deliberative situations. S-reasons capture what we wish our advisors to be concerned about, since we, the advisees, shall be able to recognize the recommendations they issue as truly helpful and non-presumptuous. This is not to reiterate the important point from above that

capturing the first person deliberative perspective is paramount for any account of reasons. Rather, it brings us to the perhaps most important source of motivation for desire-based accounts: that they, unlike most competitors, refrain from imposing reason-demands on agents that incorporate aims they have not chosen as their own; that they refrain from hoodwinking agents into compliance with alien agendas in virtue of utilizing what could be seen as nothing but a particularly sophisticated form of bluff.

The motivation in question springs from the famous *internalism requirement*. The requirement places a necessary condition on practical reasons: whether some consideration qualifies as a reason for an agent depends upon its capacity to engage what that agent cares about. Bernard Williams, utilizing semantic ascent, offers this definition in terms of statements about reasons: The statement that “*there is a reason for A to ϕ ...* implies ... that A has some motive which will be served or furthered by his ϕ -ing, and if this turns out not to be so the sentence is false.” (1981:101) Williams calls reasons that meet the requirement *internal* and those that do not *external*, but since he does not believe in external reasons, what his definition effectively does is to place a constraint on practical reasons as such. His dichotomy between internal and external reasons does not offer a classification of two kinds of reasons so much as in effect to rule out *categorical* or *external reasons*, reasons agents allegedly have irrespective of what they happen to care about. What Williams should have said is that practical reasons are internal or cease to be reasons. The requirement proposes a litmus test each reason-candidate must pass. The requirement has received extensive treatment in the literature. Here I shall be rather brief and not defend a stance for or against; all I shall provide is a rough rationale for the requirement. The employment of the requirement within my larger argument does not

rest on its validity so much as on the assumption that it figures as a vital element for desire-based accounts.

The core intuition behind the internalism requirement is that agents must be capable of seeing a point in taking seriously the demands that are applied to them. The internalism requirement incorporates a profound skepticism with regards to all forms of substantive demands and standards that are forced upon agents without their approval: dictates that they must do something even though they cannot quite see why. James Dreier captures the spirit nicely (2001:41):

When we give a justification, we are either explicitly citing or adverting to some norms. But we can't just cite any old bunch of norms. Which norms count toward justification? The problem is that if we simply cite a bunch of rules, the agent may well ask, what are those rules to me? She may ask for a reason to follow them. And we can't just shrug this off. Suppose that someone cited the laws of India in support of moral principles. We ourselves recognize that this sort of justification is useless. ... If we cite the laws of India and our subject asks what reasons she has to follow them, we understand what she's asking. She's again asking for reasons. She doesn't see any force in the rules we've cited.

It comes all down to reasons, then. And reasons, the requirement pronounces, cannot be imposed on agents from the outside as it were lest they lose all force and point. If someone, after careful reflection, were to remain entirely unmoved by some consideration, its capacity to figure as a reason is seen as systematically undermined. The question of why accept a consideration as a practical reason is always a fair one to ask. And the buck stops with agents on whom it is to answer this question. Their indifference towards a consideration, says the requirement, must mean the denunciation of it as a

reason by the only arbiters and judges there are. Considerations that have aspirations to reasonhood must thus find a foothold in the concern-structure of the relevant agents to whom they are proposed, for either they will connect to what they care about or they simply will get disconnected. If considerations fail to *engage* agents – by incorporating aims that are considered *alien* by those agents – the requirement strikes out these considerations as mere reason-pretenders falling short of qualifying as true practical reasons. Richard Joyce writes (2001:80-81):

Normative reasons claims – claims concerning what it is rational for an agent to do – must be something that potentially engage the agent to whom they are applied. This doesn't mean that the presentation of a true normative reason claim immediately results in the agent being motivated; rather, it means that the agent cannot sensibly both acknowledge that something is a normative reason for him to act and ask "But so what?" Any adequate theory of normative reasons must make out reasons to be precisely those things that forestall a "So what?" response.

How does the internalism requirement lend support to desire-based accounts? The requirement certainly does not imply the truth of desire-based accounts. Various authors, in fact, consistently endorse the requirement and yet reject desire-based accounts. The former places a necessary condition on practical reasons, whereas the latter provides a sufficient condition for reasons. The former constrains what qualifies as reasons, without necessarily saying what reasons are or to offer an account. The latter, in contrast, does say what reasons are and does offer an account. Still, how the requirement is able to lend crucial support to desire-based accounts is not hard to see. According to the requirement, principles retain normative force only if its subjects, the relevant agents, are prepared to

accept them. And accepting a principle is to exhibit some willingness to comply, which requires a motive, a desire to do something. David Velleman offers a similar reconstruction (2000:170-171):

Suppose that reasons for someone to do something must be considerations that would sway him toward doing it if he entertained them rationally. And suppose that the only considerations capable of swaying someone toward an action are those which represent it as a way of attaining something he wants ... These assumptions, taken together, seem to imply that the only considerations that can qualify as reasons for someone to act are considerations appealing to his antecedent inclinations – that is his desires or dispositions to desire.

The internalism requirement demands reason must have the capacity of engaging the concerns agents have, and desire-based accounts take care of this by precisely making reasons a function of those concerns. Desire-based accounts, hence, explain reasons in a manner that fits exactly the job description placed by the requirement. Desire-based accounts find support in the requirement because they offer the best explanation of the requirement itself. The relevant logic here is abduction, not deduction, that is to say inference to the best explanation. It is unclear whether those who accept the requirement and yet reject desire-based accounts can do equally well. Their alternative way of accommodating the requirement smacks of being ad hoc. Suppose that alternatively, reasons were grounded in concern-independent considerations, such as objective values. Dancy, among others, has argued for this, while also holding on to the requirement. The question he and his compatriots face is why, given that we are now supposing reasons spring from value rather than concerns, some value here is capable of empowering

reasons while the exact same value there is not, solely in virtue of extrinsic and unrelated facts concerning what certain agents happen to care about and what they happen to be motivated by. The question, in other words, is how to explain the relevance of such concern/motivation-related contingencies for reasons given that what reasons are, according to this alternative picture, has nothing to do with those concern/motivation-related contingencies. It does not suffice to insist that reasons must be capable of motivating agents. The question is precisely why this should be so if the source of reasons had nothing to do with motivation altogether. To date I am not aware of any satisfactory answer to this question.

The requirement, then, lends vital support to desire-based accounts. Yet the role it plays for desire-based accounts actually goes further. The requirement does not just figure as one witness among others to speak on their behalf. Without the requirement desire-based accounts are exposed defenseless with regards to the myriad of competitors who seek to empower features other than desire with reasons-related capacities. This can most dramatically be illustrated with regards to desire itself, only desires of others. According to desire-based accounts, it is not just desires as such that figure as crucial components for the reasons agents have; it is only agents' *own* desires. This is structurally somewhat puzzling. Why should only desires agents have themselves be capable of figuring in the relevant reasons-relationship and not also desires held by their agential neighbors? After all, desires are desires, and if two desires had in common everything but their location, it is not immediately obvious why their geography should make all the difference. It sounds rather awkward to say that this desire on your right can undergird someone's reasons while its identical twin desire on your left cannot. Consider

an example. Suppose Bert and Berta share a room that Berta wants to be warmer but Bert does not; Bert is in fact indifferent. Berta and Bert sit side by side with the thermometer in equal reach. Now what desire-based accounts imply is that Berta has a reason to get up from her seat and turn up the heat and that Bert has no such reason; disregarding, for them moment, any concerns Bert may have for Berta and that he also lacks any other social/moral sensitivities; imagine Bert and Berta suffer from systematic mind-blindness, a particularly heavy form of autism. How come, then, that whether the exact same desire can figure in Bert's reasons depends on its location? The answer must be that this is so because only if the desire is Bert's own do we have any guarantee that Bert will be moved by it and see a point in doing something that furthers it. In other words, because out of two otherwise identical desires the internalism requirement preempts the motivationally disconnected one to count at all.

This completes my reconstruction of desire-based accounts and of what they have going for them. Unfortunately desire-based accounts render reasons entirely void of distinctive normativity. Once we attend carefully to the details of desire-based accounts, I believe this becomes next to obvious. We must not fall prey to an awkward reverse consciousness suspecting what is readily apparent really must be deceptive instead. The crucial point is the relevant relationship between situations, agents, ends and options is blatantly non-normative or rendered thus only in virtue of externally normative principles that transcend the compass of desire-based accounts. Let me explain.

Take situation S, agent A, end E as well as options X and Y. Now suppose that as a matter of empirical fact, A will more likely realize E if he were to X rather than Y, and

everything else is truly equal. Given how A's overall cognitive-motivational system is set up, and given that he really cares about E and is not stupid, we would expect A to prefer X over Y. We would certainly expect A to gravitate towards X rather than Y and be puzzled if otherwise. The one-million-dollar question, however, is this: can we detect anything distinctively normative in that complex relation that explains why A *ought* to do X, *should* do X, *must* choose X over Y, would become *subject to warranted criticism* if he did not in fact choose X over Y, and so on? Can we detect anything that explains, based on that relationship, why preferring X over Y is genuinely the *appropriate* choice and why preferring Y over X is genuinely the *inappropriate* choice?

We cannot. What we can detect is this: if A chooses X over Y, he will more likely realize his end E. If A chooses Y over X, he will less likely realize his end E. This appears to be a simple empirical observation of the kind that psychologists, sociologists, or anthropologists might advance with regards to some study subject. It fits the perspective of some neutral observant of the strategic situation of two combatant parties: If Napoleon moves there, he will risk his right flank. If he retreats, he will compromise the benefit of the victory the day before. Granted, we may employ reasons-language to capture this and other empirical observation. This may be especially illuminating since it reveals how the relevant agents appraise the situation. It would be perfectly fine to use the vocabulary of S-reasons to do all this. Calling S-reasons *reasons*, however, does not render them reasons in any distinct normative sense, just as saying the reason why the volcano erupted was that magma was building up involves nothing normative either.

It is worthwhile to stress, then, that my present contention is not with what are often called *motivating reasons*. I am not trying here to come to grips with what could be classified first and foremost a psychological phenomenon. It is true that people usually exemplify systematic patterns in their reasoning and behavior where goals and means become suitably interconnected; and further that those cognitive/conative patterns will be subject to and evolve in response to various ecological pressures. Effective practical reasoning certainly enhances our capacity to successfully deal with our environment; it contribute to self-preservation and flourishing in a hostile world. And it is also true that with systematicity in patterns comes understanding and prediction. Given that people tend to connect means and ends, we usually are able to cite the relevant sets of beliefs and desires that illuminate why they acted as they did; beliefs and desires that effectively present their choices in a favorable light and promise to make sense of them. The issue of motivating reasons, and of what sets them apart from both purely normative and purely explanatory reasons, is tricky. But to the extent to which we can distinguish between motivating and normative reasons – or at least between two roles reasons can play, a motivational and a normative role – my concern lies with the latter and not necessarily the former. Whether our agent A is under some psychological pressure to prefer X over Y and whether he is under some normative pressure to prefer X over Y are commonly supposed distinct matters. And here I do not contest that S-reasons, or our responsiveness to S-reasons, figure as crucial components of a complex psychological capacity – to reason and plan for the future – that has enabled our species to take over the world. I am mildly optimistic about the prospects that a decently plausible psychological story shall eventually be forthcoming explaining why we exhibit these kinds of systematic patterns

in reasoning and acting linking ends and means. For all I know the story will include various evolutionary and cultural details. But this concerns questions distinct from those we are presently occupied with. My focus here entirely rests on whether agent A has distinctively normative reasons to prefer X over Y, instead of merely and fortunately being subjected to cognitive habits that have served our species well. Or, to put the question slightly differently, whether the sense in which he does have such reasons is a distinctively normative sense, as opposed to a purely motivational and psychological sense of reasons.

Returning to this question, it is safe to observe that is at least not obvious whether S-reasons are normative, and this alone spells trouble. Desire-based accounts of reasons face great opposition. The important point now is not to take sides in this debate, but notice that there are sides to take. And what opponents of desire-based accounts deny is certainly not the empirical observations from above: namely that doing X rather than Y will better serve A's end E. What they deny is that this, or this alone, determines A's normative situation. It appears, then, that in addition to whether we have the sort of complex quadruple relationship in front of us there is also the issue of whether that relationship captures what reasons are. For if there was no such further question, then specifying that relationship would be all there is to say, and the impression that there is a further issue one can potentially disagree with would amount to nothing but confusion. Yet we should not be that comfortable with passing verdict of confusion on entire traditions of philosophy.

It appears, in short, that there are two coherent hypotheses, neither one ruled out on logical or conceptual grounds alone:

(1) Given S, doing X instead of Y will better serve A's end E and (hence) S is a reason for A to X.

(2) Given S, doing X instead of Y will better serve A's end E but S is not a reason for A to X.

To put some meat on this, take some ax-murderer who wants to indulge in a particular brutal act just for the fun of it, and learns he could best achieved this by causing a blood-bath in the local kindergarten. Does he have a reason to do so, given that this is what he wants? It's controversial, to say the least, and that's all we need to focus on for the moment. Now, given that proponents of both hypotheses (1) and (2) acknowledge the fact that, given S, doing X instead of Y will better serve A's end E, there must be something else they contest, and therein precisely lies the problem for proponents of desire-based accounts. It seems they have mentioned everything there is to mention. What further fact is there in their desire-based arsenal? Since all parties agree on the empirical/relational facts that supposedly carry reasons, the question of which of (1) and (2) is true should be closed and settled, but it clearly is not.

Now, the easiest way to rule in favor of (1) over (2) is to introduce an additional principle. A principle of practical reason, rationality, or prudence that dictates that one should, given that in situation S, doing X rather than Y better serves ones ends, do X. Or perhaps it is as simple as that it is a good thing (bears truly commendatory force) if people get what they want. Kant suggested the principle that "Who wills the end, wills (so far as reason as decisive influence on his action) also the means which are indispensably necessary and in his power." (*Groundwork*, Paton translation) What

precisely figures as the best candidate for this auxiliary principle is controversial and involves tremendous complications. I shall mention a few momentarily. For now we can put all complications aside and stipulate that principle P establishes the requisite normative connection between situations, agents and ends on the one side and choices on the other. P would explain why e.g. it would be irrational for A to prefer Y instead of X given the facts are as they are supposed to be. The controversy, then, would concern principle P and its status. Unfortunately, complementing desire-based accounts with one further principle P puts enormous strains on them; it threatens to undermine what motivates such accounts in the first place.

3.5. Desire Cum Principle P

Principle P, we now suppose, has the power of settling the dispute between hypotheses (1) and (2) from above; in other words, it establishes that given agent A finds himself in situation S where doing X rather than Y better serves his ends, A has a normative reason to do X; or, at any rate, something reasonably close to that. As mentioned there are tremendous difficulties with finding an adequate formulation of P. Among them are how to control other relevant considerations; how to control relevant background conditions, such as excluders, enablers, disablers, etc; how to deal with issues relating to the scope of the relevant normative operator, in particular whether we are entitled to detach parts in instances of the overall principle. Suppose John has some end – to clean his car – which he will further by washing his car in his driveway. Do we want P to imply that John has a reason to do so? The simple solution would be yes, but

this leads to many problems. What if John's neighbor has a professional car wash station, and offered to take care of the car free of charge, and what if, in contrast, John's taking on the cleaning of his own not only involves considerable costs, but also means that he has to miss the famous one and only job interview? In such and other cases it is at least not clear whether we want to say John still has a reason to clean his car, even of the tiny-bitsy pro-tanto sort. The obvious superiority of some easily accessible alternative should bear some weight on whether pursuing the inferior option is in fact reason-supported. For otherwise reasons will proliferate in uncontrollable ways, and to point out a reason to an agent potentially will carry little recommendatory force whatsoever. A second difficulty strings from excluders and disablers. If John promised not to clean his car himself, this may exclude reasons he may otherwise have. If he has not the slightest idea of how to clean cars, this may disable the sensibility of such an undertaking. In short, the precise contours of P are complicated.

For present purposes, however, I shall put all these complications to a side. What I am concerned about is adding a normative principle to desire-based accounts in the first place, regardless of its content. The only point I need to stress, then, is that P is a *normative* principle. It comes with its own provision of oughts and shoulds which claim true normative force. Principle P is not a mere rule of thumb, a statement of some regularity, or a specification of our cognitive tendencies and expectations. Rather, it issues claims about what we should do, are required to do, or have reasons to do and so on. It introduces a practical norm the authority of which can be questioned and scrutinized. And therein precisely lies the problem.

What explains that normative status of P? If P has normative authority – if it is capable of establishing the relevant normative connections – this constitutes a most interesting normative fact in need of explanation. And now desire-based accounts face an uneasy situation. They must choose between two options: to provide an explanation or not to provide an explanation for the normative authority of P. And as I shall argue, neither choice quite works. The first choice would be certainly preferable, since desire-based accounts aspire to explain reasons and authority, and if for that purpose they invoke a principle with its own normative authority, then we want to know why they are entitled to appeal to that principle and suppose it to carry normative force. The second choice would be to suppose principle P correct and authoritative, primitively, without further explanation. This would turn desire-based accounts into versions of non-reductive realism of the sort discussed in the previous chapter. It would propose and leave unexplained one normative principle, albeit one that potentially could claim particular intuitive support. I shall comment on this second choice first.

Accepting the normative authority of P as an unexplained primitive obviously amounts to giving up on the initial aspiration of naturalism. Because of this the second choice seems a non-starter qua naturalistic candidate. Still, some patience may be in order. If desire plus principle P is the closest we can get to explaining reasons in naturalistic fashion, and if this exception is motivated, it may not be such a terrible thing to excuse one single principle from the relentless demands of explanation. After all, the principle, together with the usual set of empirical/motivational facts, promises considerable explanatory power, as every practical reason henceforth would receive a principled explanation. Unsurprisingly, then, a number of proponents of desire-based

accounts – Humeans as they usually call themselves – have urged their compatriots to accept this single and local exception from explanatory demands. James Dreier in particular has argued for this in *Humean Doubts about Categorical Imperatives*, 2001. He proposes the simple reading of principle P – or (M/E) in his idiom – such that “if you desire to ϕ and believe that by ϕ -ing, you will ϕ , then you have a reason to ϕ .” (Dreier 2001:38) As mentioned before, the simple reading probably needs further amendments. For now, however, let us ignore these difficulties and follow Dreier (2001:35) in that

According to [this] reading, it is really a normative claim. It says, in effect, that you ought to perform the necessary and sufficient means to your desired ends. You might not do this. You might, at least on occasion, find yourself lacking the motivation to perform the necessary and sufficient means to some end you desire. This would be a fault of yours, a failure of rationality. Glossing over some distinctions, we might say that your failure would be a failure of instrumental reason.

Dreier then proposes that Humeans accept principle P, or (M/E), in the form of a categorical imperative, an imperative that “a person has reason to follow it that is independent of what she desires.” (2001:37). In other words, Dreier proposes to accept the normative authority of principle P irrespective of what people desire (2001:42):

(M/E) has a kind of ground-level normative status. I think it also counts as a categorical imperative. Of course, the particular reasons that (M/E) generates are all hypothetical reasons. But (M/E) itself is not hypothetical. Its demands must be met by you, insofar as you are rational, no matter what desires you happen to have. That is why ... I think Humeans are mistaken to say there are no categorical imperatives at all.

These are astonishing words. Humeans have always dreaded the notion that there are demands that *must be met by you, insofar as you are rational, no matter what desires you happen to have*. There are many conceptions of practical rationality which involve robust and substantial demands: that one must promote one's own well-being; that one must accord weight not only on present aims but also future aims; that one must strive for flourishing or sociability or perfection. And most importantly, of course, that, on pain of deep practical inconsistency, one must respect others if one respects oneself, and surely one must respect oneself. Humeans have regarded all this with profound skepticism. What they quintessentially contest is the attempt of substantiating the authority of ambitious norms on the basis of some robust and substantial conception of practical rationality. This precisely is the characteristic Kantian move, where a failure to comply with morality, for instance, is analyzed as a failure to comply with practical rationality as such. Hence, the inception of desire-transcendent forms of practical rationality represents all that Humeans have traditionally been against. Dreier is fully aware of this singularity (2001:37):

I will argue there is something special about exactly the kind of norms of rationality Humeans accept. This special status confers a kind of necessity on the Humean norms that we may properly doubt can accrue to other sorts of norms. The request for justification, I will argue, is intelligible as a demand for reasons bearing just that kind of necessity. And we may properly doubt that the demand for moral justification can be satisfied. So we may properly doubt that moral imperatives are categorical, but we must allow that some imperatives are categorical.

Unfortunately, granting this exception involves costs for desire-based accounts that far exceed the mere admission of a gap in explanation. It amounts to no less than the rejection of the most profound rationale for desire-based accounts itself: the aforementioned internalism requirement. If the normativity of principle P is taken to be primitive – that is not to be explained in terms of desire – the amended version of desire-based accounts cum principle P inevitably places itself in stark opposition to the internalism requirement. An agent may now respond to the dictates of P, namely that he must choose the means that best promote his ends, in exactly the same manner of philosophical puzzlement as he may responds to e.g. the dictates of some moral dictate, call it principle M, namely that he must help prevent the death of a starving child. Since according to the amended version principle P claims authority over agents regardless of whether it has a foothold in those agents, there is exactly the same sort of structural gap opening up between what reason-demands are proposed to agents and which in fact are guaranteed to engage those agents; between *engaging* reasons and *non-engaging* reasons. The bottom line, then, is that one cannot consistently use the internalism requirement to fend off some other principles on grounds of their potential disconnection to desire while at the same time give principle P a free pass. If principle P is structurally exactly alike those other principle, including moral principle M, in that it too involves a claim to authority not accounted for in terms of desire, it ipso facto must share the same fate when it comes to whether it can live up to the internalism requirement.

The important point is entirely structural in character. It is also subtle, and we must ensure not to get sidetracked by various irrelevancies. This is not an exercise in psychology. As a matter of empirical fact, it would be highly unusual if agents were to

respond to the recognized ways of furthering their ends with thinking “so what, what is that to me?” Few will do so, unless they are trying to score a philosophical point. If someone were indeed to respond in that way, we would start to wonder and probably would have cause for concern. We might start counting the days he will stay around. It would be crazy for him to disregard basic means-ends rationality that secures his further existence. But then, few will respond in that way when presented with commonsensical moral principles either. If someone were indeed to respond in that way, we similarly would start to wonder, and now would have even more cause for concern: now we might start counting *our own* days to stay around. It would be no less crazy to disregard basic moral principles that secure our *coexistence*. Generally speaking, the issue is not what *is psychologically likely* to happen but what is structurally *possible* to happen. Even if it so happened that every creature in the universe accepted principle M, it still would violate the internalism requirement, because it lays claim to authority in a manner irrespective of desire. Hence, it is entirely beside the point whether principle P or some other principle such as principle M comports to a greater extent with common empirical psychological patterns. The point, plain and simple, is that by assuming the normativity of principle P to be primitive and independent of our desires, this version inherits exactly the same structural features that enabled the *engaging/non-engaging* gap to open up for other more ambitious normative domains such as morality. It is because both principles P and M lay claim to authority irrespective of desire that it becomes possible that what an agent desires and what he must do come apart. On what basis, then, are we entitled to award principle P the status of primitive authority that we are not also entitled to award to the basic moral principle M?

It shall prove most instructive if we keep alive the comparison between principle P and principle M, which represents the prototype of what the internalism requirement was devised to strike out. Consider an example of Dreier's (2001:38-39) that can easily be adjoined by a second and relevantly similar one. Consider Ann's case first.

We tell her that she ought to take a prep course for the LSATs. She asks why. We point out that she wants to raise her chances of getting into a competitive law school, and she can raise her chances by taking the prep course. She admits as much, but she still isn't motivated to take the prep course. So we cite the rule [of means-ends rationality, or principle P]. Now suppose that Ann agrees that this rule does indeed instruct her to take the prep course ..., given what she believes and desires, but she shrugs and doesn't accept the rule.

Ann obviously exhibits a severe form of practical irrationality. She would be the kind of person we are likely to get worried about. Don't plan any trips with Ann! Her practical reasoning is off, and we should fully concur with Dreier that *we must conclude that there is something wrong with Ann*. Now consider Antoinette's case:

We tell Antoinette that she ought to give her leftovers to the starving child outside. She asks why. We point out that the child will die otherwise, and that her donation involves virtually no costs to herself. She admits as much, but still isn't motivated to make the donation. So we cite the rule that one ought to help prevent a gratuitous death of an innocent child if one can do so without anything but marginal costs. Now suppose Antoinette agrees this rule indeed instructs her to help, given her situation, but she shrugs and doesn't accept the rule.

Antoinette obviously exhibits a severe form of immorality. She would be the kind of person we are certain to get worried about. Don't plan to do anything with Antoinette!

Her attitudes are repulsive, and so again *we clearly must conclude that there is something wrong with Antoinette*.

Ann, then, flouts principle P and Antoinette flouts principle M. What's the difference? The content of the two principles, of course, but nothing more. With regards to the relevant structural feature, namely the relationship of their presumed authority to desire, they are exactly on par. And since we are not presently concerned with the contents of normative principles at all, the answer should simply be: nothing, period. For sure, it is an open possibility that we conclude on substantial grounds that one principle is valid and the other is invalid. After all, they differ in what aspects of reality they enable to carry reasons; principle P places considerable weight on the concerns of agents and what furthers them while principle M places considerable weight on the capacities of agents to prevent gratuitous deaths. But this difference in content is irrelevant for present purposes and anyways, thus far we have not encountered any substantial argument for or against each principle. It is certainly not the injunction that Antoinette ought to hand over her leftovers that any participant in the present debate would want to take issue with. To paraphrase Bernard Williams, there are many things we want to say to her: "that [s]he is ungrateful, inconsiderate, hard, ..., nasty, selfish, brutal, and many other disadvantageous things. ... it would be better if [s]he were nicer..." (1995b:39) But neither do I wish to take issue with means-ends rationality. Three cheers to means-end rationality! Both principles, then, involve norms most would find intuitively compelling and few would wish to discard. The question does not concern our de-facto inclinations towards both principles, but that they both lay claim to authority in the same desire-independent

manner, and that consequently, as far as the internalism requirement is concerned, they both stand or fall together.

The amended version of desire based accounts, then, attempts to quarantine what has the capacity of giving reasons to agents – exclusively, namely, what has the right kind of relation to what they care about – only to invite back in a principle that brings with it all the features the quarantine was erected to keep out in the first place. The line the amended version seeks to draw demarcating acceptable from unacceptable principles is inherently instable: by drawing it where it wishes it to be drawn it must reemploy an element plainly falling on the other side and hence censured off hands. Yet desire-based accounts cannot have it both ways. They ultimately shall prove incapable of withstanding the dual pressure from left and right, which is simultaneously be applied by those, one the one side, who seek to acknowledge the authority of additional desire-independent principles, and those, on the other side, who even question the authority of the single principle P: Philosophers such as Hume, who famously contested the authority of means-end rationality, such as contained in principle P, when he wrote *Tis as little contrary to reason to prefer even my own acknowledg'd lesser good to my greater*. (Treatise 2.3.3.) To respond to the skeptics who insist each principle must be accounted for and who are prepared to jettison all wanting an explanation, proponents of the amended version must plead to allow one exception, and to accept principle P despite its disconnection to desire. Yet thereby they inevitably become vulnerable on their other flank.

And indeed, those who wish to advance more ambitious principles to carry normative force are quick to point their finger at precisely this instability. Christine

Korsgaard notes that "... once this kind of irrationality is allowed in the means/end case, some of the grounds for skepticism about more ambitious forms of practical reasoning will seem less compelling." (1996:321) Here, then, do we find two representative figures questioning the soundness of the amended version of desire-based accounts. Korsgaard, coming from one end, will press to allow in principle M as well, given that principle P, which is already supposed inside, no less violates the internalism requirement than does principle M. And Hume, coming from the other end, will press to disallow principle P as well, given that principle M, which is already supposed outside, no less violates the requirement than does principle P. Both Hume and Korsgaard appear consistent in a way in which the amended version does not. Desires-based accounts, then, cannot consistently refuse to accept principles such as principle M on grounds of its independence to desire and yet at the same time accept principle P even though it exhibits just the same sort of independence of desire.

Desire-based accounts, I conclude, are not entitled to accept the authority of principle P without explanation. This precludes the second option of how they could incorporate principle P. The first one, which we initially declared preferable anyway, is to provide an explanation for the authority of principle P. Do desire-based accounts, then, have the resources to offer an account of the authority of principle P? Unfortunately, they have not.

The only element desire-based accounts can draw on to explain the authority of some principle is desire. When, and in virtue of what, we like to know, does an agent have a reason to comply with a principle? The answer, desire-based accounts must say, is

when, and in virtue of the fact that, the agent has accepted the principle; when, in other words, honoring the principle features as one of his concerns. Suppose an agent accepts some principle and now faces a choice where he can either do something that honors the principle or alternatively do something which dishonors it. He must have a reason to do what honors the principle, we hear the imaginary proponent of desire-based accounts saying, because it furthers something he cares about, namely to honor that principle.

But now we squarely find ourselves in a circle. We introduced principle P precisely because we recognized that the mere empirical correlation between what someone desires and what furthers his desires cannot carry distinctively normative reasons. A gap remains, and principle P was called upon for the sole purpose of closing that gap. Hence in the attempt of explaining principle P's own authority one cannot presuppose what principle P was employed to explain in the first place, namely that with the appropriate provision of desire comes reason. In other words, if what is at stake is the reasons-related capacity of desire, it doesn't help to add yet another desire. If the transition from what furthers desires to what we have normative reasons to do requires an extra step, we make no strides by co-opting an extra desire for taking the relevant step.

Return, then, to Ann's case from above. Ann wants to get into law school and recognizes that she will more likely be able to do so if she takes a prep course for the LSATs, but then refuses to do so all the same. We propose principle P to Ann, and she acknowledges that principle P directs her to take the course under those circumstances. She still has no inclination to take the prep course. Now, we concluded there is something wrong with Ann, but wondered what exactly is missing. "What Ann is missing can't be

any desire” Dreier keenly notes. We cannot get Ann to take the course by just adding yet another desire (2001:39):

The desire that is supposed to bridge the gap between believing that a rule requires her to ϕ and being motivated to ϕ is the desire to comply with the rule. But suppose that Ann’s mental inventory were supplemented with a desire to comply with the rule, in this case to comply with (M/E). Could this complete the picture? Were she to desire to comply with (M/E), would she then be motivated to take the LSAT prep course? By hypothesis Ann ... fails to be motivated by the acknowledged means to her desired ends. So adding a desire (complying with (M/E)) does not in her bring about the motivation to perform an acknowledged means to her end of doing well in the LSAT. ... This futile attempt is exactly what we would be engaged in if we were to try to bring Ann to take the LSAT prep course by giving her a desire (complying with (M/E)) that would motivate her to take the prep course. So what Ann is missing cannot be a desire.

Dreier frames the argument in terms of motivational force, but it works just as well in terms of normative force. If Ann generally doubts she has reasons to do what furthers her ends, she will not regard her newly instilled desire to comply with (M/E) to conclude in a reason either. Supplying her with the goal of honoring principle P only entails that she now has a reason to comply with the principle if we are entitled to suppose that she has reasons to do what best satisfies her goals, which, of course, is nothing but a restatement of principle P, the very principle at issue. This addition in desire thus cannot change the situation. Previously we dealt with a single desire of Ann’s, and we were flabbergasted by her unwillingness to proceed to means she knows will implement that single desire. Now we are dealing with two desires of Ann’s, one about getting into law school and one about honoring principle P, and Ann again exhibits the

same degree of unwillingness to proceed to means she knows will implement these two desires. The number of desires, however, is not the problem; rather, the problem entirely lies in the transition from desire to reason, be it one or many, and thus we are simply moving in circles here.

Dreier calls this line of reasoning the Tortoise argument given its resemblance to the famous story by Lewis Carroll. There we encounter a clever tortoise unwilling to draw an inference of the modus ponens sort and who is adjoined by a rather dull Achilles attempting to entice her to do the same by offering one after another inference of the modus ponens sort. If the Tortoise stubbornly remains disinclined to accept modus ponens – or doubts she has reasons to do so – then presenting yet another modus ponens argument for accepting modus ponens will not make the slightest difference, neither for her inclinations nor her conception of what she has reasons to do. Likewise, if we already question that desire-based reasons contain distinctively normative force, then providing a desire-based reason on behalf of the authority of desire-based reasons cannot do the trick either. If we don't get from desire to reason, we don't get from desire to reason, period, regardless of whether the case at hand is of the concrete sort on a first-order level or involves principles on a higher-order level.

The fundamental problem, then, is not (1) that by accounting for the normative authority of principle P by way of desire we inevitably would limit the scope of its application; that only those agents who harbor the relevant class of higher-order and principle-related desires could truly be said to have reasons to implement what they acknowledge are means to their desired ends. We could come to accept that only Ann in

the second case but not in the first case has reasons to take the prep course, even though both are fully aware that taking the course increases their chances of getting what they want. Nor is it (2) that the authority of principle P would become escapable in a rather blunt manner, for all one needed to do in order to relinquish its burden is to abandon the requisite higher-order and principle-related desire. Perhaps we cannot just discard desires directly, but a good recipe for achieving this indirectly would be to violate the principle with great regularity such that its psychological hold eventually wears off. Nor is the problem (3) that it becomes hard to see how principle P could retain much independent force and point; that one would either have the requisite motivations to honor principle P, in which case one would also have reasons to do what one is inclined to do anyway; or one would not have the requisite motivations to honor principle P, in which case one would not act contrary to any reasons. The absence of the corresponding inclinations would conveniently excuse one from being held answerable to the principle. It is not immediately evident, to say the least, how anyone could ever come to *dishonor* principle P in the first place. Put together, what (1) and (2) and (3) cause are rather marginal worries by comparison. The fundamental problem is that even with respect to someone who does have the requisite principle-related desires and who does renounce all plans to escape it and who is motivated to honor it – the perfectly practically rational agent – we still have no account of why he should have any reasons to choose those means he acknowledges further his ends.

The amended version of desire-based accounts thus collapses. Accepting the authority of principle P without further explanation undermines the very rationale for desire-based accounts in the first place. And the attempt of providing an explanation of

the authority of principle P has proven futile. The principle-adjoined version of desire-based accounts does not represent progress over the principle-free version.

3.6. Value

Turn briefly to value. What I have to say about value recasts much of what I have said about reason before. Moreover, desire-based accounts of value are somewhat less complicated than their sister accounts of reasons, and their incapacity to capture the distinctively normative import of value is also more immediate.

According to desire-based accounts, value enters the world through our concerns: What is valuable is what figures as the object of our concerns. To be valuable is to stand in the appropriate relationship to those concerns; namely to matter to us, to be something we have a stake in, to be something we are involved with and are not indifferent about, and so on. Just as in the case of reasons we best understand this desire-based relative as theories of what value is, not of what provides or confers or generates value, and identify what is valuable as things out there, as something being the case, rather than as concerns getting what they want. I value professional exchange within my department. It is the sort of profitable interaction between likeminded philosophers I value, not that I shall be able to cross out one of my concerns from some list of things to worry about. Value structures our plans and projects. It directs us to do something, to make something happen. Desire-based accounts must, and I believe are able to, capture the world-directedness of value just as much as they did so with regards to the world-directness of reasons. The question, however, is whether they also are able to capture the normative status of value.

Many find it rather evident that they are not. In his book *Value Judgment*, Griffin writes (1996:23): “What is clear is that tastes and sentiments, as such, have none of the authority that we attach to values. That is, the mere fact that my feelings prompt me to approval or disapproval would, in fact, cut no ice, even with me, unless I could sort my feelings into better and worse, sound and unsound.” I believe Griffin is on target here, though his metaphor of *cutting no ice* does not yet bring it into clear focus. What he writes later on is more helpful. Introducing what he calls a *perfectly general requirement of value*, he says “For anyone to see anything as valuable, from any point of view, requires being able to see it as worth wanting.” (1996:28) The notion of *worth wanting* of course is itself normative; it brings us directly to the core of what is at stake. If value is normative, its normativity must be reflected in the value-desire relationship itself.

Value solely based on desire, however, is stripped of all distinctively normative features. Desire-based accounts lack resources to capture the normative significance of value. Value is supposed normative precisely in the sense that it is said to merit choice, to render certain responses towards it appropriate and inappropriate. Value is that which is desirable. But the meaning *able* adds to *desire* in *desirable* is importantly different from that of *audible* and *tangible*, which denotes only that which de facto can be heard or touched under normal circumstances. Yet to speak of something as desirable is not to speak of it as something that can be desired, or even as that which is commonly desired, but as something that *should* be desired, or *justifiably* is desired. Not to desire something that is desirable is in fact to commit some sort of mistake. Indifference towards recognized value is not just unusual but inappropriate; it diagnoses a fault that can justifiably be criticized. Not to hear what is audible, in contrast, is just to miss something.

At most it diagnoses only some regrettable decline in auditory capacities. A doctor may suggest measures to slow down that decline in response, but this is not to level any sort of warranted criticism. I believe it is rather apparent that the sense of *desirable* that desire-based accounts are capable of capturing is only the one akin to the audible sense and not that of the normative sense; of what is de facto and commonly desired as opposed to what is appropriately desired. All desire-based accounts can capture is some version or other of *being desired*. *Being desirable*, in contrast, must appear forever beyond their scope.

This is not to contest that there is a sense of value that neatly corresponds to objects of desire, a non-normative sense as it were. When we speak of the cash-*value* of a car, what we are after is what people are prepared to pay instead of what they ought to be prepared to pay. I believe economic theory is in fact best understood as talking solely about that non-normative kind of value and utility, that is about that which people de facto cherish, regardless of what if anything they ought to cherish. Economic utility is merely an operational notion as they say. In that respect, the scenario of monetary-value-without-desire is quite unlike that of value-without-desire. High prices usually prompt interest, but in no way mandate or warrant it. Someone's choice not to value what is economically valuable does not express any deficit on his part but merely a divergence from common patterns. And often it may even be the thing to do.

The reason why value solely based on desire lacks normativity is not that desire may not have the appropriate depth or complexity – we have seen it very likely has *that*. Our responses are usually motivated in terms of others. Rather, desire seems to be the wrong kind of foundation for value altogether; nothing in the relationship between desires

and their object points towards anything that could warrant or justify being so related. Desire may be constitutive of value; but even then would value fail to merit desire. It couldn't do so anymore than being a triangle merits being three-sided or being water merits being H₂O. In fact, as these examples suggest, the problem may even be aggravated if desire was constitutive for value. It is hard to see how desire could take credit from that which it cannot help but bring about.

Moreover, if value is normative, it seems fair to suppose that there must be cases where either desire is warranted but lacking or unwarranted but present. Yet on the constitutive model underlying our present version of desire-based accounts, neither option appears even possible. A value that would unsuccessfully call upon a corresponding desire could not even exist; value, in turn, would come all too easily to desire; desire never could be completely unwarranted. If desire does not respond to but rather create the property of value itself, the normative potential of that property vis-à-vis desire simply evaporates. How can that which fully originates in desire lend credentials to its originator? The incapability of desire-based accounts to imbue value with normative power seems thus to be a direct consequence of the ontological response dependency it postulates; it seems to make value completely epiphenomenal; desires cannot invest their objects with the sort of critical power or significance that is essential for the normativity of value. Only if values are capable of providing a substantive evaluative standpoint for the corresponding attitudes can value be said to be normative. If, in contrast, it is impossible that in light of one's values one's attitudes are ever inadequate, those values lose any normative point. Value would follow desire like shadow follows body, and

thereby provide value with no more critical power or meriting force than shadows tell walkers where to go.

3.7. Desire, Ideal and Informed

Let me briefly address the most popular response to the difficulties just presented, namely to move away from actual desires and focus instead on informed and ideal desires as the foundation for value. I shall focus mostly on value instead of reasons in this final section; partly for easy of exposition, but more importantly because the progress the ideal presumably affords is most immediately seen when it comes to value. Suppose, then, that value does not emerge on the basis of what is in fact or actually desired, but on the basis of what would be desired under ideal or more informed conditions. My claim is that moving from actual to informed desires does not solve the problem of authority. It aggravates it.

Now, to begin with, the kind of argument presented above against actual desire-based accounts of reason and value can easily be extended to ideal desire-based accounts as well. The problem there was to identify any distinctively normative character in the relation between, as in the case of reasons, what's desired and what furthers it; or between, as in the case of values, what's desired and, well, what's desired. Recall the argument vis-à-vis reasons. As before in the case of actual desires, we are still facing two competing hypotheses in the case of ideal desires. The only difference is that now we are dealing with *ideal ends* instead of *actual ends*:

(1') Given S, doing X instead of Y will better serve A's ideal end E and (hence) S is a reason for A to X.

(2') Given S, doing X instead of Y will better serve A's ideal end E but S is not a reason for A to X.

And again, given that neither (1') nor (2') is ruled out on logical or conceptual grounds, we can ask what determines the truth or falsity of (1') and (2'), respectively, where it cannot be the mere presence of the quadruple relation involving situations and agents and ideal ends and choices, which is countenanced by both hypotheses. The source of the problem lies in the relation, not in their parts or relata, and hence replacing one item for another does not solve that problem. I shall not further pursue this line of argument, however. Instead, what I shall focus on in the remainder is what allegedly presents itself as a clear instance of progress in the ideal version, and argue why this perceived advantage is in fact a disadvantage.

What is the supposed advantage? Now turn to value. In the actual version, we found a perfect match between what is valuable and what is desired, and consequently found no basis on which desire could ever become merited or justified. What is supposed to supply that basis, value, turned out the mere slave of desire, with no say and independent critical force. What we missed, in other words, was any sort of guidance for what ends to adopt: what ends are truly choice-worthy and not just chosen; what is apt for admiration and not just admired; what is fitting for reproach and not just reproached; and so on. It is at this juncture where the informed version promises the looked-for remedy. What we actually desire and what we ideally desire represent separate domains. As a

consequence, it becomes possible that the former finds guidance in the latter. What we do desire can fail to match what we would desire under ideal circumstances.

The crucial question, however, is whether that mismatch itself has normative potency. When we not desire what we would ideally desire, does that constitute a failure that warrants criticism, or does it merely constitute a divergence between two competing desire-perspectives? Consider the following case. You and I have different desires about what I should desire; you would like me to care more about your financial hardships, having an urge to turn over half my salary, while I do not want myself to have such generous attitudes. That I depart from what you want that I desire is no failure, but merely a divergence. In order to retain any normative force, our attitudes towards our ideal desires must not be like that. Unfortunately, as we shall see, that's exactly what they must be like. Once we know what ideal desires are we must loose all interest in them altogether.

What are informed desires, then? The informed version just as the actual version represents a classic response dependent account of value. It explains the value of states of affairs in terms of desire. Moreover, the explanation it offers aspires to be a complete explanation. Having the appropriate informed desires in place entails there being value; there being value entails the appropriate informed desires being in place. The informed version must consequently refuse to explain or motivate *any* desire *ultimately* in terms of the value of certain states of affairs. The qualification "ultimately" is important, since just as other response dependent accounts it may explain and motivate certain responses in terms of others. An agent's desire to have an X may be motivated by some of its valuable

properties P, so long as the value of these properties in turn is explained by some other responses towards P. Saying that *X is desired* may thus not be the end to the story about its value. Yet somewhere the story needs to end, and according to ideal desire-based accounts, it has to end with some informed desire rather than some value.

There is a choice we face between two alternative models of informed desires. Using some handy terminology introduced by Michael Smith, I will call one the ‘example model’ and the other the ‘advice model.’ (*Internal Reasons*, in: *Ethics and the A Priori*, Cambridge 2004). Since the second is clearly preferable, I only mention the example model to motivate the advice model. According to the example model, the desires that confer value are those an agent would converge on if he underwent a process called ‘cognitive psychotherapy,’ which proceeds as follows. Start with what the agent in fact desires; update and revise what the agent believes; change in beliefs will causally bring about change in desires. Continue the process until a stable and suitably enlightened state is reached where no further update in belief would result in any additional change in desire. The desires resulting from such a hypothetical process would *set the example* for what is valuable for the actual agent. (Cf. Brandt 1979).

The reason why the example model is unsatisfactory is simple. What is valuable to agents must adequately reflect their actual situation, including its limitations, and not supplant their actual environment by some ideal utopia. Consider knowledge. Naturally, the fully informed counterpart of the actual agent would harbor no desires for taking measures to acquire further knowledge. She already has it in buckets. For example, she would not want to finish that book since she already knows the ending. Yet coming to

know more might be supremely valuable for the *actual* agent. Finishing the book may be the most valuable way to spend her time. Since what is valuable for the actual agent must be sensitive to the actual situation in which she finds herself – including her epistemological condition – we need another model. Such is provided by the advice model. According to this model, the desires that confer value are those the ideal version would want her actual version to have. (Cf. Railton 2003) Even if the ideal version knows the books ending, she may still advise her actual version to finish it. Taking full account of her actual situation, the ideal version would fix her actual version's values in terms of the advice she has to offer.

What remains to be settled, then, is how to specify *the advice* in the advice model. What exactly is it that my ideal version wants my actual version to want; how does his superior epistemological position determine the contents of his desires about what I shall desire? This is no easy question; at least not for non-instrumental desires, of which there must be plenty; instrumental desires are partially conditioned by factual information, and so there are no principal obstacles for understanding how belief revision may result in revision of instrumental desires; yet for non-instrumental desires the problem becomes rather intriguing, mostly for purely formal reasons. Belief purports to represent how things are, while desire only how things should be; in familiar parlance, belief and desire differ in their direction of fit, which right there undermines any hope for formally valid inference rules relating belief and desire. Inferential relationships are usually understood in terms of truth, such that the truth or likelihood thereof of one proposition is affected by the truth of other proposition(s); yet because desires aren't amenable to truth or falsity, desires consequently aren't amenable to being objects of inferences either. Thus, if an

update in belief is to settle what to desire, it must do so in other than formal or inferential ways.

It is most important to emphasize, then, that we must not think of the relevant advice in terms of instrumental advice. This we already have, for what provides reasons according to actual accounts is what *in fact* best furthers our ends; what is valuable is what *in fact* stands in the relevant relation to our concerns. The solution to the problem of instrumental misinformation is already contained in the very way in which the actual desire-based versions are set up: the relevant relationship is specified not in terms of what agents may or may not know and be mistaken about but plainly in terms of what is the case. We already have on board the advisor who honors our actual concerns and merely helps us finding better ways of implementing them. Hence, the advice the informed version offers better not be what we already have, only at far less a cost, namely the cleaning up of instrumentally misinformed desires. What furthers mistaken instrumental desires never was empowered with reasons and value related capacities, for it does not deliver and sometimes does even compromise the non-instrumental desires on which the former are contingent. Cases where an agent wants to drink some glass of liquid that he mistakenly thinks is gin rather than gasoline or where he wants to quench his thirst by drinking a glass of milk that in fact would aggravate his dehydration present no serious challenge to actual versions. What those agents want is to have a glass of gin and to quench their thirst; and plainly enough, drinking gasoline or milk will not get them what they want.

Besides instrumental advice what other ways of offering advice are there, then? In our discussion of the example model, we already saw one possible answer: update in belief may settle what to desire in a purely *causal* manner; there we find one solution for how belief may determine what to desire, if only in principle, which is all we are looking for anyway. The causal account thus presents *a* workable solution we may always fall back on, which is indeed what I propose to do momentarily. Still, I cannot wish to embrace a causal account simply as an unmotivated background assumption, for I know many defenders of informed versions will not be entirely comfortable with it. So I'd need to motivate it, and here is how: if we are unable to find alternative solutions for how update in belief may determine what to desire, we have no choice but to (partially) resort to the causal account as the only account we know to work. This is not to deny that those alternative solutions may make some contribution; they may very well limit the range of desires to be adopted in light of greater knowledge. Still, if after applying all available non-causal procedures the question of what to desire remains open, the causal story inevitably must emerge as at least part of the mix. By considering the most plausible non-causal accounts and finding them insufficient, I wish to put back the ball into the camp of defenders of informed desire-based accounts. If they think update in belief can completely determine what to desire in other than causal ways, please say how. Until then, I have to do with what I know to work.

One popular yet minimal constraint on how my better informed version may select which desires of mine to sanctify is this: only those, namely, which have a realistic chance of being satisfied. The underlying rationale would be this: *desires aim at the attainable*, and unattainable desires are thus disqualified. If, unlike me, my better

informed version knows certain desires of mine to be unattainable, he may discard them. This constraint is a minimal one, for it certainly applies only to a tiny subset of what I may or may not desire; it certainly cannot be the only criterion. Still, many seem to find this rationale plausible beyond question; even Hume, who thought that no desire could be contrary to reason, conceded that desires may be deemed unreasonable in case their objects do not exist, desires that is which are in principle unsatisfiable. I wish to object, however. Many desires make perfect sense even though they aim at what we know cannot be realized, or what we know to be unattainable. I may desire to live longer than I possibly can as the biological creature I am; or, to take another example, Nomy Arpaly may at times desire to be Bertrand Russell – an admittedly unlikely superhero she wants to be. (Nomy Arpaly 2006) Once one sees the pattern, one can easily generate additional examples; fathers who want to be at their son's football game here and at the job interview there at the same time, spouses that want to ask their deceased loved ones for advice, or sport fans that would like to have been around for a famous game to watch in person before they were alive. Now the interesting twist in all these examples is that we do know already that those desires are unattainable; and yet, we do not see anything wrong with retaining those desires because of that. Yes, I know I can't live a thousand years, but I really and strongly would like it so. My response to the acknowledged unattainability is not to relinquish that desire – even if I could – but rather to regret its unattainability; that unattainability is “what I'm taking issue with,” not the desire itself. The same is true for Nomy Arpaly. She of course knows that what she desires cannot be, given minimally controversial principles of modal logic and personal identity; that's why she chose those examples in the first place! Yet all the same, Nomy Arpaly feels no

inclination to discharge or wish to discharge that desire for this reason. In what sense are we mistaken or irrational? After all, desires are not “chosen” in response to what we think can or cannot be, but in response to what we’d like to be: that I live longer than I biologically can, and Nomy Arpaly be Bertrand Russell. Now, given that both of us do not seem to lack any crucial information and still hold on to our desires, what would the basis be for my even better informed version to jettison that desire? What is it that he knows but I do not that would warrant the removal? I cannot see what.

Next, consider coherence; my ideal version may choose what desires to sanctify in response to what would make for a maximally coherent desire set. Now, to put an initial worry upfront, many authors have forcefully questioned the soundness of the very notion of coherent desires. (Cf. Sayre-McCord 1997) It remains unclear how to apply a notion developed for representational states to non-representational states in such a way that would preserve the disqualifying nature of incoherence. We know why incoherent beliefs are tainted; taken together, incoherent beliefs depict the world in ways it cannot be, and beliefs aim at truthful depiction. Yet what is the analogue case against incoherent desires? If desires are said incoherent if they cannot be jointly satisfied, it is hard to see what’s wrong with incoherent desires. What is wrong, for instance, with our father’s dual desire to be at his son’s game and at the interview at the same time? Granted, incoherent plans and intentions may be bad, for what they specify to be done cannot be done in the world as we know it; plans that cannot be carried out plainly enough are bad plans. Desires, however, have no such immediate impact on what is to be done, as they enter plans and intentions only indirectly and in conjunction with other states, some of them presumably representational in character.

Now, though I find this line of argument convincing, let me argue a somewhat weaker claim which should suffice for my overall motivation of the causal account. Even if the notion of coherent desires was sound, considerations of coherence cannot fully settle how updated beliefs settle what to desire, and hence need to be supplemented. Supposing that my actual desire set is less than fully coherent, there are a billion different ways of making it so. Imagine different developmental paths representing each way in which my ideal version may want to revise my desires such as to make them more coherent; since each path equally leads to a perfectly coherent desire set, he cannot determine which one to choose solely based on considerations of coherence. In the face of such vast possibilities for how he may make my desire set more coherent, considerations of coherence thus seem to offer little guidance. This is not surprising; just consider the analogous case of belief. If I was told that only those beliefs of mine are rational that I would still uphold if I made my beliefs as a whole more coherent, I would have little clue as to which those are; small sets of beliefs already generate vast numbers of different ways in which I may make those sets more coherent; suppose I have only two incoherent beliefs, I may restore coherence by discharging either one of them, or both, or adopt additional beliefs that would clean up the incoherence, etc. In light of any realistic estimate of how many beliefs or desires we have, the problem at hand truly becomes intractable if the only consideration in play was coherence.

So we need additional considerations. Perhaps my ideal version could decide which desires of mine to sanctify on the basis of their (indirect) contribution to my well-being or happiness; most probably, my ideal version will have concerns for my well-fare, and so perhaps he might determine what he wants me to desire on the basis of those

concerns. I'm skeptical this could go very far. First of all, we need to make sure that this proposal doesn't smuggle in what we are seeking to explain, namely value and well-being. The suggestion is non-circular only if we understand the relevant notions of happiness and well-being employed here not already in value-laden terms. It simply won't do to say that my ideal version would select those desires which would increase my well-being.

Now, trying now to interpret the proposal charitably and non-circularly, it still faces considerable difficulties. While I think that it is true in some cases – particularly drastic ones such as the devastating craving a rejected lovers harbors for his love — that considerations of how certain desires might (indirectly) impact my happiness may be helpful, I doubt that such considerations get us very far in general. For one thing, in the face of the vast number of different ways in which desires can be chosen with an eye to happiness, the proposal generates more indeterminacy than it resolves. Also, it encapsulates a rather trivial and unhelpful element. If the idea is to select desires on the basis that if I had those desires I would take pleasure in them or their objects, this would give my ideal version little guidance indeed. It is true for virtually all desires that if I had them I would take some pleasure in their satisfaction. Personally, I'm not a big sports person; and though I'm aware that I would take great pleasure in sports if I was more passionate about sports, this leaves me utterly cold; realizing this comes on the cheap. The fact that if I desired X, X would bring me some pleasure simply doesn't tell in favor of any particular X, for one may substitute X for whatever one pleases.

But more importantly, I think the suggestion at hand plainly is confused about the nature of desire itself. It betrays the notion of desire by giving it an overly and implausible hedonistic rendering; in fact, desires need not be and often are not connected to one's well-being or happiness at all; construing them so often misses their very point. What we desire is simply that certain states of affairs obtain; not necessarily that we receive happiness from seeing those states of affairs to obtain, though of course we may desire this as well – we do desire happiness, after all. Desire-satisfaction accounts of value are commonly thought superior over hedonism exactly because they make it intelligible how things could be of value to people even if those things have no immediate psychological impact on them. Recall Griffin's example of the caring father: "... if a father wants his children to be happy, what he wants, what is valuable to him, is a state of the world, not a state of his mind; merely deluding him into thinking that his children flourish, therefore, does not give him what he values." (Griffin 1986:13) Examples such as Griffin's are quite common; suppose I desire – or would want me to desire if fully informed – that the next world chess champion be Swiss. What would be valuable in this case is *that the next world chess champion be Swiss* and not *that I take pleasure or satisfaction in the next world chess champion being Swiss*. The nationality of the next world chess champion would or would not have value independently of whether I will come to know the results or whether I would take pleasure in the results. Again, the object of the relevant desire is simply that certain states of affairs obtain; if they do obtain, value will be realized, regardless of how or even whether it will affect me psychologically. What matters according to all desire-based accounts of value is *de re satisfaction* and **not** *felt satisfaction*. Hence, questions of what to desire cannot

straightforwardly be settled in terms of what would make one happy or what would contribute to ones *felt satisfaction*, lest we fundamentally misunderstand the nature of desire.

Now let me put all this together and generalize the points from above. Take two states of affairs – *Fa* and *Gb*, and suppose my ideal version needs to decide whether he wants me to desire *Fa* or *Gb*; suppose further that this is a mandatory yet exclusionary choice, and everything else is truly equal: each desire would make the same contribution to how well off I will be, or how much pleasure I would receive in *Fa* in case I desired *Fa* or in *Gb* in case I desired *Gb*; each desire would (indirectly) give my life equal point and worth; and each desire would equally well fit in with what else I happen to desire; both would have equal effects on my deliberative efficiency and instrumental rationality; using a locution of Michael Smith's, the addition of each desire would equally well preserve a "maximally coherent desiderative profile." Now, as we have seen from above, my stipulation of everything else being truly equal should not present principal problems, given the vast number of different pathways for how conative profiles may be developed. Yet because everything is truly equal, considerations of coherence, instrumental rationality and contribution to felt satisfaction cannot settle what desire to adopt. In what way then can we assume that the superior epistemological position of my ideal version settles this question? We certainly do not want to say that my ideal version simply affords greater insight into matters of value; this would clearly defy the response-dependency the ideal version is still committed to – to explain value, not to presuppose it. But if it is not greater insight into value, what else will decide whether my ideal me wants my actual me to desire *Fa* or *Gb* instead?

The only informative answer I'm aware of again proceeds in broadly causal-procedural terms. It starts out with the idea that *my* ideal version is still a version of *me*, albeit ideal. The advice I receive does not come from strangers, ideal observers, or even God; it is importantly still my own advice. In order to make good of this idea, we again start out with what I in fact desire. We then update and revise my beliefs; change in beliefs will causally bring about change in desires. We continue the process until a stable and suitably enlightened state is reached where no further update in belief would result in any additional change in desire. The outcome will be an ideal version of myself with second-order desires regarding what I in my actual situation shall desire.

I thus take the advice model and the example model to converge on the *process* that generates informed desires. It is certainly true that the two models differ profoundly in the *structures* and *contents* of the desires they generate; the advice model places all emphasis on indirect and second order desires, whereas the example model is first order and direct; only in the example model but not in the advice model need the relevant ideal desires reflect what the ideal version wants for himself or cares about; but with respect to the way in which the relevant informed desires are "generated" or specified I take both models to be roughly the same. They depart from actual desires and causally modify them in light of greater factual knowledge. If an advocate of the advice model finds this an uneasy concession to make, she owns us a substantial alternative explanation. To this day I know none.

But now it should be evident why the desires we would want ourselves to have under those conditions cannot offer any guidance for our concerns after all. With regards

to our actual concerns we know why the goals it accords value should engage us: They concern what we actually care about! Yet this is not so when it comes to our ideal concerns which are but causally altered concerns. The desires of our idealized version are not ours, and consequently *de re* satisfaction of informed desires is not the same as *de re* satisfaction of what we actually care about. The greater the divergence between what we actually desire and what we would want us to desire under idealized conditions, the more pressing the question becomes why it matters for us that our idealized rather than our actual desires are satisfied. It would certainly matter if we had such idealized desires, but given that we don't, why should we care about them?

To give it more bite, I'd like to present this thought experiment. Suppose your idealized version of yourself manages to write you a letter across possible worlds. In this letter he or she recommends you to pursue a list of goals. The letter starts: "I'm a completely different person now. The way I see things now has no resemblance to how I saw them before, which is as you see them now. What I previously found important appears now to be vain. What previously appeared to be vain I find now important. I know you won't understand. I am very sorry to be unable to explain my reasons to you, which go beyond what you can grasp. None of my explanations would make any sense to you. P.S.: Attached is a list of what you should desire from now on." As you quickly come to realize, the listed goals are diametrically opposed to what you care about most. While you aspire to become an academic, the letter recommends becoming a farmer. While you desire to have a family, it recommends a purely solitary life. While you seek a career in a middle sized western city, it recommends that you settle to some unknown

village far away. Now you start to wonder what reason you could possibly have to revise thus drastically your own most intimate goals.

Chapter 4

Against Agency-Based Accounts

4.1. Introduction

This chapter examines the hypothesis that authority is grounded in what is constitutive of agency and action. *Agency-based accounts* of normativity have recently received a great deal of attention, and are characterized by Luca Ferrero's as follows (2009:304):

[T]heir basic claim is that the norms and requirements of practical rationality and morality can be derived from the constitutive features of agency. Hence, a systematic failure to be guided by these requirements amounts to a loss of agency. But there is a sense in which we *cannot but be* agents. It follows that we are necessarily bound by the oughts of rationality and morality, we are bound by them *sans phrase*.

The locution *sans phrase* is used to contrast the manner in which agency-constitutive norms are supposed to bind from norms that only bind according and in relation to yet some other norms. Agency-constitutive norms are supposed authoritative not just according-to-agency-constitutive-norms themselves or according-to-yet-some-other-set-of-norms, but simpliciter, *sans phrase*.

Agency-based accounts represent a Kantian alternative to the previously discussed non-reductive and naturalistic forms of normative realism. Bernard Williams writes (1995:174-5):

Consider another picture of what it would be for a demand to be 'objectively valid'. It is Kant's own picture. According to this, a demand will be inescapable in the required sense if it is one that a rational agent must accept if he is to be a rational agent. It is, to use one of Kant's favorite metaphors, *self-addressed* by any rational agent.

The Kantian notion of norms self-addressed by and to agents, respecting their deliberative autonomy, looms large in agency based accounts. Kantians reject as incompatible with autonomy the supposition that demands can be legitimately imposed on us from the outside, demands legislated *for us* instead of *by us*. The complaint is that this would circumvent our agency in a way that would betray the very normative question in the first place. Proponents of the agency-based accounts regard the normative question as essentially an agential question: why shall *I* take seriously what demands are posed to me? It is essential for the question that it is addressed to agents on whom it is to answer it. The force of the question flows from the potential disconnect between what may be asked of someone and what he shall be able to reflectively stand by. That is the problem.

The hope is “[t]hat a clear statement of the problem is also a statement of the solution.” Contemplating the authority of morality, but intending the point to apply more generally, Korsgaard continues: “If the problem is that morality might not survive reflection, then the solution is that it might.” (1996:49) Agency-based accounts have a clear conception of what the normative question is, a mark that distinguishes them from many competitors, a conception that also is supposed to contain the answer. Once an agent has determined to wholeheartedly stand behind some principle, its normative standing is established. Furthermore, if the agent has no choice but to stand behind the principle – because it enables his very agency which generates the question in the first place – the gap between a principle and its reflective endorsement motivating the normative question has been closed. The normative question is answered once and for all.

Connie Rosati develops the idea further. In a remarkable essay, Rosati juxtaposes the normative question to Moore’s open question argument. Rosati focuses on the normative dimension of personal value, but what she says can easily be generalized to other dimensions as well. The reason why she believes many traditional analyses of what is good for us retain an open feel is that they do not yet take seriously enough our autonomous agency. She writes, quoted here at length (2003:496-7):

I shall try to show that Moore’s argument directs us to a genuine, and little appreciated, problem for naturalism ... I hope thereby to explain, at least in part, the evident and, to many, perhaps mystifying vitality of Moore’s argument. In making out my claims, I will urge an understanding of the argument that is different from Moore’s. As Moore presents his challenge, the open question argument poses a theoretical question to be asked and answered from the standpoint of the ethicist who tests proposed analyses

of good in an effort to arrive at a scientific ethics. In contrast, on my understanding, the argument reveals something of genuine theoretical importance about value while also exposing a practical question that most naturally arises from the standpoint of the agent who is inquiring about what to value. As Moore concludes his challenge, the argument has supposedly provided a vehicle for undermining all forms of naturalism. In contrast, on my understanding, the argument may not defeat naturalism so much as set it a task. If my suggestions are on the right track, what the argument really undermines is neither naturalistic accounts of good nor analyses of good, as Moore originally thought but, rather, accounts that do not fit with our agency.

Continuing a few pages later (2003:505-7):

I believe that for the argument to have force, we would have to be agents, creatures with the capacity to engage in autonomous evaluation and action. The force of the open question argument has deep roots in our agency; it is not merely a function of the expressive and recommending functions of our evaluative notions. And so ... a successful account of personal good will have to fit us as agents. ... We seek to explain why we have not yet reflectively converged in our theorizing about personal good on a particular N. A plausible explanation would point to something about the type of creature we are that enables us rationally to question the normative authority of any N proposed thus far. It is our capacity to evaluate and act accordingly that enables us to be self-governing—that is, to decline to follow the press of some natural desires and impulses while endorsing or at least reflectively acceding to the press of others. That capacity would seem to be engaged when we question proposed forms of naturalism and find them wanting: we question the natural standards they give us, much as we question our own desires and impulses, identifying considerations that count against treating them as normative. When a proposed account fails to pass muster relative to our autonomous

evaluation, when we find ourselves wondering whether things with N really are good for us, it is as if we were detecting a mismatch. My proposal is that the mismatch is with the very features that enable us to question whether things with N are good for us: those features that make us agents. The kind of account of personal good that would escape such questioning would be one that could survive our reflective scrutiny, that would fit with our capacity for autonomous evaluation and action.

The basic tenets of agency-based accounts are easy to state. The issue of a norm's authority is taken to be characterized in relation to the agential question of why take seriously what it demands. A privileged few norms at best shall have the capacity of answering that question in the positive. Which and why? Agency-based accounts propose to award the distinction of authority to norms only in case they are constitutive of the very essence of who we are as deliberating and autonomous agents.

The fusion between descriptive and normative elements, between a characterization of who we are and a prescription of what we should do and ought to do, is fully intended. Constitutive principles are introduced to elegantly bridge truths about who we are with truths about what we are supposed to do, thereby dispelling whatever qualms one might have harbored about normativity. Korsgaard puts all this together as follows: (2008:7-8):

I must say what I mean by a constitutive principle. First, what I will here call a constitutive *standard* ... is one that arises from the very nature of the object or activity to which it applies. It belongs to the nature of the object or activity that it both ought to meet, and in a sense is trying to meet, that standard. Constitutive standards apply most obviously to objects that have some standard use or function or purpose. ... Constitutive standards are

opposed to external standards, which mention desiderata for an object that are not essential to its being the kind of thing that it is. ... Two things are important to notice about standards of this kind. First of all, constitutive standards are at once normative and descriptive. They are descriptive because an object must meet them, or at least aspire to meet them, in order to be what it is. And they are normative because an object to which they apply can fail to meet them, at least to some extent, and is subject to criticism if it does not. This double nature finds expression in the fact that we can criticize such objects either by saying that they are poor objects of their kind ... or by saying that they are not such objects at all ... Second, constitutive standards meet challenges to their normativity with ease: someone who asks why a house should have to be waterproof, or an encyclopedia should record the truth, shows that he just doesn't understand what these objects are for, and therefore, since they are functional objects, what they are.

Constitutive standards are praised capable of maintaining authoritative standing without external help, without embroiling us in any dubious normative metaphysics, and without the need to recruit forces beyond its own boundaries, and that's why Korsgaard believes "constitutive standards meet skeptical challenges to their authority with ease." (2009:29) Whether that shall be so easy, however, and whether all in all the picture is as rosy as suggested is open to some serious doubt.

And indeed, even at a first cursory glance agency-based accounts are overshadowed by problems. For starters, the marriage between the descriptive and the normative is plagued by internal conflict and dissonance. How can a principle constitute a phenomenon and yet allow that divergences from the principle count as normatively

deficient instances of the phenomenon as opposed to no instances of the phenomenon altogether? Korsgaard is well aware of the problem (2008:9):

Sometimes people are puzzled by the idea that you can fail to conform to a constitutive principle – if following the principle is constitutive of the activity, and you fail to conform to it, then aren't you failing to engage in the activity after all? In one sense that is right, but in another it cannot be, for if you were not engaging in the activity after all, then your failure to conform to its constitutive principle would not be a failure at all.

Another problem agency-based accounts face is to identify norms as necessary for the exercise of agency as such, and necessity claims are hard to establish and easy to assault. The chapter examines in detail the two most prominent proposals, advanced by the philosophers David Velleman and Christine Korsgaard, and the reasons why neither is successful cannot easily be summarized in brief. But there is a general problem underwriting both the contours of which can be stated in brief. If norms possess authority when their systematic violation undermines a creature's agency-related capacities, it follows that each and every creature navigating under the banner of agency must exhibit suitable conformity with the norm. All identifiable agents, in other words, real or imaginary, must exhibit a suitable degree of respect for the allegedly constitutive norm. And there are a lot of agents on whom one can test that claim. The result is not encouraging. Virtually all conceivable norms appealing to our humanity and rationality consistently meet agents who intransigently disrespect their prerogatives. Whatever norms are suggested as candidates for being constitutive of agency immediately invite massive assault by counterexample, a game philosophers know how to play, steadily

forcing agency-based accounts to retreat to more general and weaker norms in order to sustain the necessity claim characteristic of constitution.

This, however, puts in serious doubt how much *normative fertility*, using a phrase of Lucca Ferrero's, the constitutive programme is able to retain. To withstand the pressure into some normative wasteland, where only the minutest set of norms retains a claim to authority, proponents of agency-based accounts ultimately have little choice but admitting that constitution cannot be the entire story about normative authority, perhaps not even the most interesting part thereof. If so, the constitutive approach at best offers negligible guidance for the practical quest of figuring out how to live, what standards to adopt and abide by, and what person to be, since the range of systems of norms that, individually and collectively, is consistent with what is constitutive of agency is virtually without limits. Yet apart from this limitation in practical guidance, the modification in explanatory ambition also threatens the internal stability of the constitutive programme, or so we shall see.

Worse still, the retreat to the acknowledgment of being only part of the story itself does not work. The hope is that the constitutive programme could explore a particularly interesting corner of the normative geography, even if not its entirety, for constitutive norms are allegedly unique in that they hold authoritative for agents no matter what, since they are inescapable. The hope shall be frustrated too. Inescapability and authority represent rather distinct phenomena and to equate them is to equate apples and oranges, as the final section of the chapter demonstrates.

Parts of the chapter are reminiscent of the transcendental strategy discussed in chapter two, in particular with regard to the connection between the inescapability and authority. Since that strategy was dealt with in chapter two, this chapter does not elaborate any further.

4.2. Methodological Preliminaries – What Constitutes What?

Constitutive norms are a common phenomenon, showing up in countless practices, most easily discernible in games. Games incorporate rules that constitute what they are. Though not without problems, I shall not expound on any potential philosophical difficulties affixing the notion of constitutive principles as such. I am content with using John Searle's old-fashioned distinction between regulative and constitutive rules (1969:33): "regulative rules regulate antecedently or independently existing forms of behavior; ... constitutive rules do not merely regulate, they create or define new forms of behavior." According to Searle, regulative rules have the form "Do X" and "Do not do Y if you care about Z," whereas constitutive rules have the form "X counts as Y in context C." That's all I'm going to say about the form of constitutive principles.

Constitutive relations hold between some constitutive base and some constituted target. The question is in virtue of what do the kinds of constitutive relations hold that agency-based accounts are after? Where do we have to look for an explanation of why certain principles (the base) supposedly are constitutive of being an agent (the target)? Two broad options present themselves. First, the relation might hold in virtue of the

structure of certain natural kinds. Second, it might hold in virtue of certain conceptual truths. Water is constitutively identical to H_2O , and bachelors are constitutively unmarried. Thus, we can ask: does the constitutive relation agency-based accounts postulate between the implementation of certain norms and agency hold as a matter of natural constellations or conceptual truth? As we shall see, neither choice is without problems, yet under scrutiny the first holds up worse than the second.

Consider first constitution in virtue of the structural compositions of natural kinds. Water molecules are constituted by H_2O and watery stuff is made up of lots of water molecules. Chemistry is the exemplary science that studies constitutive relations in structural compositions. If you want to know what some natural kind is made out of, you need to take it apart and expose its structural organization. Unfortunately, this approach fares poorly as a schema along which to unpack the constitutive claim of the agency-based programme. One reason why is that agency is a very unlikely candidate of a natural kind with clear-cut boundaries susceptible to structural decomposition and chemical analysis. The methodological upshot of this is that we have no independent criterion allowing us to antecedently identify agency as a natural kind that we could subsequently examine under the microscope. We cannot just pick it out by pointing at it. In order to identify agency we already face all the substantive difficulties that a natural-kind approach could ever hope to answer. How much rationality is required for agency? How much control and what kind of control? How much cross-temporal continuation and internal integration? Do agents have to wholeheartedly self-identify, and if yes what does that mean? In other words, the methodological approach to agency as a natural kind would presuppose a solution to the problem it is supposed to solve in the first place.

There are no independent boundaries out there delineating agency that could be consulted to answer our questions. For that reason alone the natural kind approach appears hopeless.

Yet even if there was a solution to these methodological problems, there is a more fundamental and immediate reason why the natural kind approach is ill suited for the constitutive programme. The natural kind approach is of no normative and meta-normative consequence whatsoever. Suppose there was a natural kind of agency which was constitutively underwritten by the implementation of certain norms. What normative and meta-normative relevance could that possibly have? The relevant question would be why be an agent, a member of that particular natural kind? Yet there is no more normative pull for almost-yet-not-quite-agents to become agents as there is for almost-yet-not-quite-water-molecules to become water-molecules. In case science uncovered the nature of agency – still supposing, contrary to fact, that there is a natural kind of agency – one could, out of curiosity, query whether oneself happens to be an agent, an inquiry not entirely unlike some recent research into the human genome trying to figure out whether, for instance, it contains Neanderthalean traces, which it apparently does, indicating some degree of prehistoric cross-species mating. And should the result concerning one's own agency come out negative, one would have no cause to be troubled whatsoever. One would simply not be an agent, who cares? It is hard to detect even a hint of a normative implication contained in the natural-kind approach that could possibly explain why one should strive to become, and why one should ensure to stay in case one already is, a member of this peculiar natural kind. As far as I am aware of, no proponent of agency-based accounts has ever adopted this approach.

Consider, then, constitution in virtue of conceptual truths. We know many constitutive relations are established conceptually. Libraries are constituted by collections of books, universities by buildings, students and teachers, museums by collections of art and a constant flow of visitors. This is simply what we mean by *libraries*, *universities*, and *museums*. Or consider an example from the realm of classical music, namely different orchestra constitutions. Why does a philharmonic orchestra have a larger constitutive body than a chamber orchestra? The answer lies in the ways the categories *philharmonic orchestra* and *chamber orchestra* are used by professional musicians. They conceptualize different compositions of musicians found apt for performing different styles of music. The tradition of classical music and its corresponding classificatory schemes explain why philharmonic orchestras require a constitutive body of around a hundred musicians while chamber orchestras require a smaller body of only around fifty musicians. End of story. If an apprentice were to search for some deeper metaphysical truths explaining why the New York Philharmonic was constitutively composed of around a hundred players, we would have to redirect his focus away from metaphysics to the prevalent classificatory practices within the tradition of classical music.

All this is pretty straightforward and it hardly surprises that most proponents of agency-based accounts have explicitly interpreted the relevant constitutive relationship in conceptual terms. Thus understood, we would say that the implementation of certain norms is constitutive of agency and we would explain this constitutive relationship by appealing to the ways we have come to use the concept *agency*. Perhaps the concept was introduced in order to draw distinctions that matter to our larger interpretative endeavors. Since, according to this second conceptual approach, we are not supposing that the

category *agency* tracks a natural kind, a certain stipulative character in the conceptual approach shall be unavoidable, with conceptual analysis ultimately revealing how certain communities tend to classify. The basic idea in the conceptual approach is, then, that it is in virtue of conceptual truths – our classificatory practices – that *agency* bears a constitutive relationship to certain norms, just as it was the case in the example of *philharmonic* and *chamber* orchestras.

4.3. Agency, Shmagency

Unfortunately the conceptual approach is overshadowed by problems as well, and major parts of the chapter are dedicated to exposing them more thoroughly and in connection with the various concrete constitutive proposals. The devil famously is in the details, which is nowhere truer than in the case of agency-based accounts. Still, the problem has a general structure which I'd like to lay bare before I turn to my piecemeal critique of the problem as it surfaces in the particular proposals. The problem is the *Agency-Shmagency* problem, and David Enoch has led the charge in a remarkable piece of the same name in the *Philosophical Review*.

Here's a rough outline of the problem. Suppose that as a conceptual matter, abiding by norm-complex *N* turns out to be constitutive of *agency*. And, suppose further, this is so because the concept of *agency* has been set up to categorize systems that abide by *N*. We are supposing, then, that this is simply what *agency* means, just as what *philharmonic orchestra* means is an ensemble constituted by a body of at least a hundred musicians. Now if you are abiding by *N* you are an agent and if not then you are not. But

how could that piece of information provide you with any guidance as to whether to abide by N? Why is it relevant for the normative and meta-normative status on N? From the deliberative standpoint, what the conceptual findings about *agency* furnish is but an alternative way of putting the very same deliberative question: whether you should strive to be an agent, which is but another way of asking whether to be a system that abides by norm-complex N.

Enoch puts the charge beautifully and I'm going to quote him at length. He illustrates the problem in connection with the constitutive account of Christine Korsgaard's, forecasting many of the crucial details and their difficulties occupying us later. Still, I believe the passage and the argument it articulates are fully intelligible in their own right (2006:178-9):

[C]onsider Korsgaard's hope of grounding a reply to the skeptic in what is constitutive of action. We are to imagine, then, someone who remains indifferent when we tell him that his actions are immoral or irrational. He then reads Korsgaard and is convinced that self-constitution is a constitutive aim of action, so that you cannot even count as an agent and your bodily movements cannot even count as actions unless you aim at self-constitution of the kind Korsgaard has in mind. And assume that our skeptic is even convinced that – miraculously – morality and indeed the whole of practical rationality can be extracted from the aim of self-constitution. Do we have any reason to believe that now he will care about the immorality or irrationality of his actions? Why isn't he entitled to respond along the following lines: "Classify my bodily movements and indeed me as you like. Perhaps I cannot be classified as an agent without aiming to constitute myself. But why should I be an agent? Perhaps I can't act without aiming at self-constitution, but why should I act? If your

reasoning works, this just shows that I don't care about agency and action. I am perfectly happy being a shmagent – a nonagent who is very similar to agents but who lacks the aim (constitutive of agency but not of shmagency) of self-constitution. I am perfectly happy performing shmactions – nonaction events that are very similar to actions but that lack the aim (constitutive of actions but not of shmactions) of self-constitution. ... So what is it to me how you classify my project?"

The challenge Enoch poses to proponents of conceptual agency-based accounts is this: Define agency as you like, and stipulate whatever host of norms you wish to be constitutive of agency. Why should that give you any title to recommend these norms to your audience as authoritative on the mere basis of the relevant definition of agency? Why should your audience not suppose that all this conceptual maneuver achieves is to equip them with an additional way of asking the normative question, namely why to be an agent, as opposed to be a shmagent or to be no XX-gent at all? David Lewis ridiculed the approach in similar ways, quoted at the beginning of Enoch's essay (2006:169):

Why care about objective value or ethical reality? The sanction is that if you do not, your inner states will fail to deserve folk-theoretical names. Not a threat that will strike terror into the hearts of the wicket!

And of course, the crucial point is not that folk-theoretic names are unable to strike terror because their corresponding sanctions are minute or because of the case-hardened nature of those to whom they deny their application. Instead they lack critical normative force altogether. What lies at the heart of practical normative criticism is not a charge of misclassification. Moral condemnation does not express a linguistic misstep, and moral competency does not consist in being literate in a particular language game.

For sure, conceptual stipulations frequently carry instrumental significance. They are embedded in larger practices set up so as to differentially respond to them. Thereby they can come to be associated with substantial social and emotional pressures. If everyone is to collect a reward for meeting the socially defined standards of agency, we may want to take some notice of the corresponding stipulations after all. Or those conditioned to feel distress when denied the label may incur an incentive to earn the label for the sake of their emotional health. Yet these ways clearly are nothing but distractions to our meta-normative quest. Everything can be associated with anything and thereby import all sorts of connotations. What remains true is that how we chose to label and classify ourselves and our actions are not relevant factors by themselves that could settle what to do and accept. Classificatory schemes can only possess significance derivatively, as a proxy for *what* they classify and *its* significance to us. They trace what matters but cannot matter themselves. Representational schemes and stipulations offer no substitute for facing up to the substantive questions about normative significance.

The general conclusion I'd like to draw from this is that the conceptual alone cannot carry the normative significance we are after and which is sought to be explained by the constitutive programme. Conceptual truths are not of the right kind to speak to the normative question. As a consequence, if constitutive relations were sustained by conceptual routes alone their normative contribution would be undercut as well. It would all come down to whether or not someone deserves certain labels, which in turn would depend on whether he meets the required conditions classifying along whether or not one honors certain norms. And yet labeling issues are not what we are concerned about in our normative and meta-normative investigations.

I said *I'd like to draw this conclusion* by way of qualification, since I harbor little confidence that the rather quick remarks of this section could establish so grand a conclusion. Being cognizant of this, the best I can do is to document the force of this general charge in detail and in connection with the concrete agency-based accounts that have been proposed. Doing so shall occupy the better parts of the remaining chapter, only afterwards turning to the most prominent reply to Agency-Shmagency, namely that agency is inescapable, and that derivatively the norms constitutive of it are inescapable as well, on the basis of which we supposedly are entitled to infer their authority. This way of organizing the chapter permits a more thorough motivation of the inescapability reply itself, identifying it as a move the success of which is crucial for the constitutive program, thereby setting the scene to bury the reply with greater effect. But first, turn to constitution and belief, the example proponents of agency-based accounts love to use in order to illustrate their approach, and see how this best case scenario holds up under scrutiny.

4.4. The Aim of Belief

If there is one application fitting for the constitutive programme it is the case of belief. According to a prominent proposal belief is said to aim at truth or to track the truth. And aiming and tracking quintessentially are norm-involving phenomena. Moreover, if aiming and tracking together with their underwriting norms represent the constitutive feature of belief, the question of why those norms hold authority for belief may be answered. The promise is that we may derive the norms governing belief from

nothing more than a clear-eyed recognition of what belief essentially is. With aims come conditions for success and failure, and that is precisely what norms are all about. The question of what norms legitimately apply to belief could be settled by an appreciation of what it is those norms are applied to in the first place. The answer to the question *why believe only what's true* would reside in the very same question containing the concept *belief*. David Owens states the promise after raising the following question: "Why should believers be held to these norms? Where does their authority over us come from? Call this the problem of authority. If believing involves having an aim and these norms can be seen as instructions about how to achieve that aim, the solution is obvious." (2003:284) Not quite so obvious, I shall argue. If an explanation of authority is what the constitutive programme is expected to deliver we shall be disappointed. The central argument of this section will be that even if the constitutive programme does work for belief, what it explains is not authority but something else, something a position such as mine has no troubles countenancing. In fact, as later chapters will bring out more fully, it may even play into the hands of that position.

This section will be organized around a fascinating and recent exchange between David Velleman and Nishih Shah, which commenced in various essays by Velleman which were subsequently critiqued by Shah and which ultimately cumulated in a co-authored reconciliatory piece. The discussion lead to the most developed constitutive theory of belief I am aware of, and I would like to retrace parts of that development first before I turn to my critical analysis. In particular, I'd like to bring out why a purely regulatory account does not satisfy the normative ambitions of the constitutive account, and why as a consequence norms need to be involved more directly. With norms being

brought to central stage we can then turn to our meta-normative question of what status those norms must have in case the constitutive account turns out correct.

Begin, then, with our slogan that belief constitutively aims at truth. So familiar rings that slogan that its metaphoric character almost goes unnoticed. A more specific account is needed. To begin with, it is agents who have aims, not beliefs, and even though agents may subject activities to aims and thereby lend them aims derivatively, few beliefs are intentionally formed by their believers with any aim in mind. Most beliefs, especially perceptual beliefs, approach agents without conscious invitation or purpose. “Truth must be the aim of belief,” Velleman writes, “but it need not be an aim on the part of the believer; it may instead be an aim implicit in some parts of his cognitive architecture.” (2000:19) To spell out the metaphor while appreciating that a great number of beliefs are not formed with agential aims in mind we need to focus on the cognitive mechanisms that regulate the formation of beliefs. What we want to know is what principles are to govern the mechanisms of belief-formation so as to make good on the proverb that belief aims at truth.²

What the metaphor expresses is a special regard for truth that is characteristic of belief. To appreciate more precisely how belief relates to truth we must distinguish it

² Notice that agential-aims and state-aims can come apart in funny ways. Suppose someone wishes to momentarily deceive himself. Suppose he needs to form a false belief on some matter (fill in your favorite story why). The belief, being a belief, supposedly still aims at truth, even though the aim towards which the agent intends to form that belief is falsity. Appreciating this intriguing possibility need not unduly embroil us in the intricate matter of doxastic voluntarism. It is just the recognition that states can have inherent aims and yet be subjected to aims contrary to them by their agents. Owens suggests the case of a liar, who intentionally issues false assertions; still, assertions supposedly aim at representing only what’s true. Only under that assumption, or something close enough, is lying in fact possible. (Cf. Owens, 2003:285)

both from the conative attitudes such as hope and desire, as well as from belief's cognitive brethren such as assuming, supposing and imagining. Velleman reminds us that the unique role truth plays for belief is more subtle than suggested by the platitude that to believe a proposition is to believe it to be true. Similar things can be said of all propositional attitudes, which all bear some relation to truth. To imagine or to assume a proposition is to imagine or assume it to be true, and to desire or to hope a proposition is to desire or to hope it to be true. We need to push further. In particular, we need to appreciate the difference between the various forms of cognitive attitudes which is less straightforward yet more instructive for our purposes than the difference between the cognitive attitudes and the conative attitudes. The cognitive and conative differ in their direction of fit, in that they take their propositional object as true rather than as to be made true. To believe, as well as to assume or to imagine, is to regard the relevant proposition as true rather than as to be made true. What all cognitive attitudes share, in Velleman's phrase, is their *acceptance* of the proposition involved. What is the difference, however, between belief and the other species of the cognitive attitudes, in particular what is the difference between acceptance-in-the-belief-way and acceptance-in-the-imagining-, supposing-, or assuming-way?

Velleman's proposal points toward the special and immediate authority truth holds over belief which it does not also hold for other cognitive attitudes. What distinguishes belief from the other attitudes is that *whether to believe p* is decisively and exclusively settled by *whether p*, whereas *whether to imagine, to assume* or to *suppose p* is not decisively and exclusively settled by *whether p*. In fact, *whether to imagine, to assume* or to *suppose p* may have very little to do with *whether p*. I will follow common

usage and call this feature of belief *transparency*. Belief is transparent to truth in a way in which imagination and assumption is not (2005:499):

The feature that we call transparency is this: The deliberative question *whether to believe that p* inevitably gives way to the factual question *whether p*, because the answer to the latter question will determine the answer to the former. That is, the only way to answer the question *whether to believe that p* is to answer the question *whether p*. By contrast, the answer to the question *whether p* will not settle either the question *whether to suppose that p* nor the question *whether to imagine that p*, and so those questions do not give way to it—or, as we shall say, are not transparent to it.

The task is to capture transparency,³ and the thesis that “believing involves regarding a proposition as true with the aim of so regarding only if it really is” (Aim:251)⁴ is supposed to do precisely that. As already mentioned, however, the metaphor of aiming has to be cashed out in terms of certain (partly sub-personal) mechanisms that regulate the cognitive attitude of regarding true only what really is true or only what is appropriately supported by evidence⁵, rather than what serves particular

³ There might be more work for characterizing belief. Belief differs not only from suppositions and imaginations, but also from guessing (Owens 2003) and from what Tamar Gendler has worked out under the description of *alief* (Gendler 2008). Yet guessing as well seems to be an acceptance aiming at truth. And *alief* too is concerned with truth.

⁴ The task of capturing transparency is also related to the intriguing *wrong kind of reasons* problem. It seems to me that a successful account of transparency could also be used as an account of why certain considerations are ill suited to favor beliefs even though in some sense they show it to be a good thing to have the relevant beliefs. That I’ve read a claim on some anonymous webpage might not be a good reason to believe it, but at least it is of the right kind. That believing the claim would make me feel good is not just a bad reason, but of the wrong kind altogether. Cf. Pamela Hieronymi 2005.

⁵ There are some questions concerning the relation between truth and evidence, and it would mean for the role evidence has for belief should belief constitutively aim at truth. I am aware the connection is not as straightforward as it may initially appear. Still, in this section I’m going to assume that truth and evidence

argumentative or dialectical purposes, as it might be the case with assuming and supposing. This brings us to the regulatory account, introduced by Velleman as follows (2000:17):

Belief aims at truth in the normative sense only because it aims at the truth descriptively, in the sense that it is constitutively regulated by mechanisms designed to ensure that that it is true. ... Belief thus aims at the truth in the same sense that the circulation aims to supply body tissue with nutrients and oxygen.

The regulatory account posits (some sub-personal) belief-regulating mechanisms that are cashed out in terms of certain counterfactuals, broadly akin to Robert Nozick's famous truth-tracking account of knowledge. Nozick developed his proposal as a response to Gettier's counterexamples to the traditional analysis of knowledge, where he took belief as a primitive. Nozick argued that to know a proposition is to believe a truth which one would not have believed had it been false and yet which one would still have believed in some alternative scenarios where much is changed but where the truth of the original proposition is left intact. Still, elements of Nozick's counterfactual analysis of knowledge could be applied to the case of belief-regulating mechanisms. If, unlike assuming and imagining, believing is to retain a special sensitivity for truth that reveals itself in its characteristic regulatory patterns then some changes in truth must be systematically correlated to some changes in belief, all of which would have to be expressed by some suitable set of counterfactuals.

are systematically related, and that the implementation of the constitutive goal truth for belief must involve a special sensitivity for evidence.

Unfortunately the regulative proposal does not work. Or, to be more accurate, it does not work as an account that delivers on the idea that belief has an aim with all its intended normative implications, including transparency. Now it is important to recognize that not all authors writing on the subject share the conviction that belief is essentially normative (Papineau, Davidson), and for them a purely regulative account may be all in order. Here I cannot engage the extensive literature on whether indeed belief is essentially normative, and the criticism I now wish to advance only pertains to conceptions of belief that embrace its normative nature. Now, as to why the regulative account has no normative upshot, the short answer is simple. It is one thing to think of beliefs as states that are constitutively regulated by mechanisms bearing certain counterfactual relations to truth. It is a matter entirely different to think of beliefs as states that aim at truth in the normatively fecund sense that they would fail to do what they are supposed to do in those cases where they diverge from the common regulatory pattern. As we shall see momentarily, it may not even be possible for belief to depart from its constitutively regulatory patterns, and the problem this generates is quite profound. The important observation now concerns the way we would have to conceptualize divergent cases *were they possible*: Divergence from common patterns does not amount to a normative shortcoming. The unusual and exceptional is not the same as the deficient and degenerate.

Shah and Velleman acknowledge as much. Once the notion of belief having aims is understood in terms of regulatory mechanisms, (2005:499):

the claim that belief's standard of correctness can be derived from its constitutive aim becomes harder to sustain. That claim would now require

the assumption that if an attitude is regulated by a truth-tracking mechanism, then it is correct if and only if true. And this last assumption, though previously defended by one of the authors, has subsequently been abandoned in the face of objections from the other, to the effect that how an attitude *ought* to turn out is not necessarily determined by how it *is* regulated.

Now it is true that if the mechanisms regulating belief were to have the *function* or *design* to deliver truths, the notion of belief aiming at truth could possibly be regained. We could do so in virtue of the uncontroversial and heavy-duty normative component that the notions of function and design bring with them. The distinctive mark of functions is that they can *malfunction*, and consequently, if belief is part of some mechanism that has the *function* to regulate its states by an exclusive concern for truth then clearly beliefs fail in that regard – they *malfunction* – should they not be so regulated. Velleman shifts between different formulations of the *belief-aims-at-truth-thesis*, and tellingly in one passage he states: “My thesis is that belief is an acceptance regulated in ways *designed* to ensure that its content is true.” (Aim:277, emphasis added) But if that’s the account then what we would like to learn is how the belief-regulating mechanisms come to have functions or designs at all, and in particular how they come to have the functions with this precise content, an explanation that is not provided solely by their actual regulatory patterns. For sure, the workings of our belief-forming mechanisms confer some crucial benefits on their possessors and have done so for a long time, a fact that partly accounts for their etiology. It is fair to observe, however, that the search for a normative account in our evolutionary history has proven rather difficult if not entirely unsuccessful, though I cannot further pursue this question here. In any case, nothing Velleman says supports the

design assumption, support very much needed if it were to bear the entire normative weight.

The upshot, in rough terms, is that a functionalist analysis of belief – one that stipulates characteristic input/output patterns for belief – does not capture the normative dimension of belief, should belief indeed have such a dimension. Purely regulatory or functionalist accounts resemble the characterization of certain chemical processes, and so we might briefly illustrate the present point with an example from chemistry. Consider the simple process of polymerization, where, with the help of various catalysts, individual monomers hook up with each other to form polymers, through a process where electrons are being freed up enabling individual monomers to connect into a chain. Suppose principles P constitutively describe the process of polymerization. This example enables us to illuminate, very quickly, the two fatal shortcomings of regulatory accounts qua normative accounts. The first is this. Should a chemical process deviate from principles P it would thereby fail to qualify as polymerization, and consequently, it would thereby fail to qualify as bad or deviant polymerization. There never could be deviant polymerization in the first place. The second point is this. Even if, per impossibile, there could be deviant polymerization, the sense of deviant here would not be normative, but merely statistical. The chemical process would be deviant in the sense that it is departing from common patterns, not in the sense that it failed to do what it was supposed to do. There is *no supposed to do* whatsoever contained in the purely regulatory.

This completes the short answer as to why the purely regulatory account does not work, and it is related to the longer one to the same effect that Velleman and Shah provide (2005:500):

An adequate account of belief must explain not only the fact that truth occupies the sole focus of attention in doxastic deliberation but also the fact that evidentially insensitive processes, such as wishful thinking, occasionally influence belief. An account of belief rendered solely in terms of truth-regulation may be able to explain one or the other of these facts, depending on how the term ‘truth-regulation’ is interpreted, but no single interpretation of ‘truth-regulation’ can explain both facts at once. Hence, an adequate account of belief must include more than the fact of its being regulated for truth.

According to Shah and Velleman regulatory accounts face a dilemma. There is a stronger and a weaker version of how to construe truth-regulation, and neither quite works.

Consider the stronger version first. Here the idea is that unless an attitude is regulated by a concern for truth and nothing but the truth, the attitude simply doesn’t qualify as a belief in the first place. This option is austere, and the problem is it is too austere. The strong version is incapable of acknowledging that there are any non-truth related factors ever to influence the formation of belief. For as soon as we find some factor unrelated to truth exerting ever so slightly an influence on some attitude it would cease to pass the belief-test. There is a lot of research that seeks to unravel the real genesis of many of our ‘beliefs,’ with frequently disquieting results that seem to contradict our confidence that belief-formation is exclusively truth-concerned. The work

on the infamous confirmation bias and wishful thinking come to mind. Yet according to the strong version this kind of research would be internally incoherent as it is engaged in the vain attempt to better understand an impossible phenomenon. Beliefs could not be subjected to systematic biases, where what we want to believe or what would mesh with our preconceived worldview impacts the genesis of belief. What we would have to say in such instances is not that some beliefs were unduly influenced by considerations other than truth, but rather that what we are dealing with is another kind of attitude altogether. The project of cognitive psychology that seeks to reveal how frequently and thoroughly biases and interests influence belief-formation would be a confused endeavor. This is not how we conceive of beliefs.

But apart from a concern for the venerability of cognitive psychology, the strong version has unacceptable normative implications for an account that embraces and seeks to explain the normative character of belief. It renders impossible that there are bad and defective beliefs, beliefs that fail to live up to standards they are supposed to meet. This is a devastating result, and somewhat ironic too, since the primary objective is to establish normative criteria for belief, to explain why one should believe only what's true and to demonstrate how there could be more and less successful ways of managing ones epistemic household. What the strong version effectively does is to conceptually disallow the possibility of unsuccessful ways of managing belief, for it already burdens the very concept *belief* with those success criteria the very point of which was to chastise epistemic outliers that fail to meet them even though they should. The inevitable result is that the set of good beliefs coextends with the set of beliefs proper, where no space is left for offenders, who ironically would exempt themselves from being held to epistemic

standards at the precise moment where they chose to disrespect them. No criteria are left that beliefs could violate while retaining their status as beliefs, a status that is required to make epistemic criticism appropriate in the first place. Out the window go belief-guidance and belief-criticism. One could never violate norms for belief, a point that is made by many authors. Glüer, K & Wikforss write, speaking of *internal connections* instead of *constitutive regulations*, but targeting pretty much the same idea (2009:31-70):

However, to say that beliefs stand in various internal connections to one another is not to say that these connections are *normative*. On the contrary, precisely because the connections are internal or analytically necessary, they are not normative, not optional. If the connection were merely normative, it would be *possible* to violate the norm in question. That is, it would be possible to be in the one state without being in the other. This is precisely what is *impossible* if a relation between the states is internal. If the relation is internal, there is, so to speak, not enough room for any norm to enter between the two states.

Thus, the stronger version does not work, and thus we need to consider the weaker version. Here the idea is that if an attitude is, possibly among other factors, regulated by a concern for truth, the attitude qualifies as a belief. The condition, presumably both necessary and sufficient, would be that an attitude must exhibit enough relation to truth in order to qualify as belief, where it is understood that the condition does not rule out that other factors are in play as well. This option is permissive, and the problem is it is too permissive. The weak version is incapable of explaining transparency. What we would have are mechanisms comprising various factors impacting the genesis of belief, where no single factor is distinguished as the crucial one. As Shah and Velleman explain (2005:500-1):

If belief can be influenced by evidentially irrelevant processes such as wishful thinking, then its responsiveness to evidence must be weak enough to leave room for such additional influences. And if this interpretation of truth-regulation is correct (as we believe), then the manner in which belief is regulated for truth can't be cited to explain the role of truth in doxastic deliberation. For when one deliberates *whether to believe that p* , this question not only gives way to the question *whether p* but does so to the exclusion of any other, competing question, such as whether p would be in one's interest. Yet if belief were required only to be weakly regulated for truth, then the potential outcome of deliberation could be envisioned as a belief that p so long as p 's truth were treated as relevant to that outcome, without necessarily being treated as having absolute priority over opposing considerations: the question *whether p* would not have to crowd out competing, nonepistemic questions.

This completes the negative discussion of why a normative conception of belief must go beyond a purely regulative account. The positive solution Shah and Velleman offer is this (2005:501):

When one deliberates whether to have an attitude conceived as a belief that p , one deliberates about an attitude to which one already applies the standard of being correct if and only if p is true, and so one is already committed to consider it with an eye exclusively to whether p . When one deliberates whether to have an attitude conceived as an assumption or fantasy, one does not yet apply any particular standard to it, and so one does not yet have any commitment as to how one will go about considering it. This explanation of transparency leaves room for the possibility that beliefs can be influenced by non-evidential considerations, because it entails that one is forced to apply the standard of correctness only in situations in which one exercises the concept of belief. Not all belief forming processes require the subject to deploy the concept, and the

norm of truth that controls doxastic deliberation needn't control other processes. Our explanation of transparency thus allows for the fact that passions can influence belief.

Shah's and Velleman's central claim is "that conceiving of an attitude as a belief, rather than an assumption or an instance of imagining, entails conceiving of it as an acceptance that is regulated for truth, while also applying to it the standard of being correct if and only if it is true." (2005:497). The most striking feature in the development of this account is this: The constitutive norm for belief, namely that belief is correct if and only if true, is not to be derived from some prior aim inherent in belief, as was suggested before where it was held that which standards of correctness apply to belief somehow flows from what belief constitutively aims at. Rather, the fact that the norm is taken to underwrite belief is what (partly) explains why belief constitutively aims at truth. Belief is the distinctive attitude it is precisely in virtue of being subjected to that norm. I am not intending to point this out in order to set the stage for introducing a major problem. Pressing the view to the admission that it must construe belief as constitutively norm-governed is precisely where I wanted to end up. The important point is only that the teleology of belief is not explanatorily prior to the relevant norms, and in particular cannot independently specify which norms must be taken to constitute belief, but is entirely on par with it, as two sides of the same coin. The somewhat stipulative nature of the claim that truth-regarding norms constitute belief, already implicit in the chosen conceptual approach, now lies in the open.

Velleman's and Shah's final account is hybrid and gives rise to many fascinating questions in its own right. In particular, one might want to hear more about how precisely

the formation of beliefs interacts with the relevant truth-regarding standards; how is it that those standards are applied to beliefs or processes of belief-formation. If agents have to apply them to states or deliberative engagements, we potentially witness the recurrence of a problem Velleman grappled with earlier, namely the recognition that many beliefs are not agentially but rather sub-personally managed. If however the relevant application of those norms is cast in terms of sup-personal mechanisms, it is hard to see how the problems that plagued the purely regulative account are to be avoided. For now let us put aside these questions, and finally turn to the ultimate question: What would Velleman's and Shah's account show us with regard to the meta-normative status of the norms constitutive of belief, should their account be on the right tracks? Have they presented an account that would explain why those constitutive norms are authoritative for us?

The answer is negative, and the reason why is instructive and revealing. Before we turn to the argument why, however, two caveats need to be made explicit. First, even if, as I'm going to argue, constitutive norms for belief turn out incapable of delivering authority, this by itself need not present a problem for constitutive accounts as such. For constitutive accounts do not necessarily encompass the objective of explaining authority in the first place. I suspect, in fact, that Velleman and Shah never had any intention of applying their theory to that purpose. As far as I can tell, based on extensive study of his work as well as conversation, Velleman is no less a meta-normative skeptic than I am. My argument henceforth is not leveled as a criticism against constitutive account as such, but only against the further claim, for now without explicit authorship yet not without interest, that constitution explains authority. Second, as I cannot emphasize enough, my argument does not deny the presence of norms pertaining to belief, and in particular does

not deny that there are standards of correctness applying to belief. Quite the contrary, the departing assumption precisely is that not only do certain norms apply to belief but do apply *constitutively* so, and whatever norms and standards of correctness conceptually underwrite *belief* – the working assumption is that they are truth-related, but alternative conceptions face the same fate – the argument shall not contest the presence of the relevant norms. The argument even grants that we might speak of reasons to believe only what's true *according-to-the-relevant-norms-that-define-belief*, or that belief is correct only when true *according-to-the-relevant-norms-that-define-belief*. The question is whether the constitutive account can explain why the relevant norms have authoritative status in virtue of the fact that they are constitutive of belief.

My argument that it cannot explain authority has the following schema. Consider the question of *why respect the norms constituting belief* as a rough approximation of the meta-normative question concerning the status of those norms. There are roughly two ways in which this question about the status of the relevant belief-constituting norms might arise. One may think of these two modes as *internal* and *external* to the practice of believing. The first way of understanding the question is trivial, where the relevant question tautologically answers itself, with the consequence that this way of raising the question cannot be taken to speak to the issue of normative authority, which always must be construed as a substantial matter. This leads us to a second and alternative way of raising the meta-normative question. But there we shall see that the constitutive account does not even attempt to offer anything that would answer it or explain authority. The remainder of this section is dedicated to filling in the details of this argumentative schema.

Take, then, as a given that believing is descriptively and normatively regulated by an exclusive concern for truth. Suppose we can sum up the constitutive norms for believing in N, where as a matter of conceptual truth believing constitutively consists in respecting or being guided by N. Consider then the first version of our question: Why should one respect or be guided by N *in managing ones beliefs*? In light of the just stated conceptual assumption where belief bears a constitutive relationship to N our question immediately provides its own answer, and consequently appears somewhat ill-stated. This is rendered evident once we reformulate the question, replacing the concept of belief with our conceptual assumption. Here is what we get in that case: Why, in the formation of states that are constitutively N-regulated, should that formation be N-regulated? Or, why, in managing N-respecting states, should one respect N? Or, why, in the formation of states that are what they are in virtue of their exclusive concern for N should one be concerned about N? The point is simple: Once the concept of belief is included in our question, where it is agreed that the concept constitutively enjoins respect for N, the question of why respect N in forming beliefs contains its own answer.

The impression this engenders is that one is missing something. But one is not. One either is in the business of believing; a business constitutively characterized by the special regard for truth, and in that case it is a given that one already exhibits the characteristic and exclusive regard for truth. Or one is not in that business, and in that case one does not exhibit the characteristic and exclusive regard for truth. Or, one already exhibits the characteristic and exclusive regard for truth, and in that case one is a believer; or one does not exhibit the characteristic and exclusive regard for truth, and in that case one is not a believer. The question of why respect N when being in the business

of believing misses the crucial point that believing and respecting N go hand in hand, that they consequently cannot come apart, as if one of the options on the table was to believe while disrespecting N, and as if some rationale was needed to rule out that option. The question is conceptually closed, and no substantive issue lingers in the background. There is no open question argument about the status of the norms underwriting belief thus understood. It's not so much that a full appreciation of the normative underpinnings of the concept of belief provides an answer to the contested issue of why one must believe only what's true. Rather, it explains why such a question does not really arise. The very question is somewhat deceiving, making it appear as though there is an issue at stake where there is in fact none. It's like asking why one should touch all four bases when trying to hit a home run, as if there were two things when in fact there is only one for which we only happen to have two different labels.

The point of having offered various alternative reformulations of the same question, each leading to the same trivial result, is to make plain that what we find in the constitutive proposal is not so much a solution to our meta-normative problematic but rather something of a dissolution for the case at hand; thereby displaying the inapplicability of the very problematic to the case of believing so described. What initially appeared a question in search of some substantive answer now reveals itself, upon conceptual clarification, as laying bare a mere tautology. To believe, under that description, is to be solely concerned with truth. Why believe only what's true thus transforms into the question of why be solely concerned with what's true when being solely concerned with what's true (viz. when engaged in belief-formation). There is no issue the question thus understood brings out about which one could adopt competing

meta-normative positions, being a realist or anti-realist or expressivist or cognitivists or whatever. No position can be said to have the upper hand on this or to take any credit from this, because the meta-normative issue is inapplicable here.

Consider the infamous case of faith in divinity sustained despite blatant lack of evidence. Many authors have expressed especially high esteem for faith sustained despite all available counterevidence to the contrary, where it is considered an especially praiseworthy virtue to have overcome the vast sources of doubt and dared to believe. Consider a person of faith who takes a divergent and domain-specific attitude towards belief, such that he goes about regulating his “beliefs” in light of the evidence when it comes to the natural world but not so when it comes to religious matters where he permits non-evidential considerations to reign. There he dares to take the famous Kierkegaardian leap into faith. We disapprove, and say that what he is doing is misguided, pointing to our conceptual findings, and yet he remains unmoved, insisting to keep doing what he was doing. *You cannot go about believing that way*, we say, and add that *if you do anyway, what you are undertaking is not really believing but rather something else*. But this is not a particularly powerful response. All we can show him, assuming we are correct in our constitutive analysis, is that he is confused about his mental economy, that he must be aspiring to a rather different attitude, which he has every right to do, only that presently he has misclassified what he is doing when he conceives of it as an instance of believing. He cannot do what he thinks he is doing under that description. He takes himself to be laboring under the concept of *belief* when in fact he is laboring under some slightly different concept, *celief* say, where *celief* is an acceptance regulated by truth and evidence when it comes to natural affairs but non-evidential factors when it comes to

religious affairs. Since we have agreed to reserve the label “belief” for a specific norm-enabled activity, he must either abandon his classification as belief, or he must abandon his “beliefs,” which ain’t really beliefs properly speaking. It’s not that he had some items in his belief-box that we show he must remove. What we show is that these items never really were in his belief-box to begin with. He is simply conceptually confused.

What the case of the person of faith is reminiscent of is the problem initially noticed with the purely regulative account, namely its hyper-escapability: Should a standard be constitutive of some activity, then one abandons the activity by abandoning the standard, and we lose our grip on the crucial moment where one could truly still be considered to engage in the activity while disrespecting its constitutive standard. Again we find ourselves in danger of having no space left for systematically defective ways of going about believing, ways that fail to live up to standards they are supposed to meet, and we again would find ourselves incapable of explaining how there could be more and less (sic!) successful ways of managing ones epistemic household, where no criteria are left that *believing* could violate qua believing. To cite Velleman, “a subject who lacked a disposition to follow indications of truth would not be capable of forming beliefs.” (2009:133) And we could continue: since the subject is incapable of forming beliefs, he is also incapable of forming bad beliefs and consequently incapable of attracting epistemic criticism.

I mentioned this problem only in passing, for as sizeable as it is, is not quite the one I just was pursuing, which concerned the implicitly tautological nature of the question that the constitutive account is being taken to answer in this first version. To this

I now return. The problem is that what the constitutive account delivers has the wrong form to speak to our substantive meta-normative problematic. To bring out this more fully, consider another engagement which comes with its own constitutive standards. Think of the concept of a *criminal master mind*. Suppose the concept contains the proviso that no one qualifies as a criminal master mind unless he feels obliged to exhibit special diligence in executing his schemes. Now consider this question: Why should a criminal master mind, qua criminal master mind, exhibit special diligence in executing his schemes? The question seems pointless, especially after it was just stipulated that this kind of special diligence is a necessary component of criminal master minds. The reformulation again makes this plain. Why should one exhibit special care in executing ones schemes when being engaged in an activity characterized by exhibiting special care in executing ones schemes? Examples along similar lines are easily multiplied, all showing the same result: the observation that some standard is internal to some activity does not speak to the substantive question of whether the standard is authoritative.

And yet, isn't there a substantial question in the neighborhood? Indeed there is, which brings us to the second version of understanding our question *why respect the norms constituting belief*. As for the precise way of understanding it we can already take a clue from our last example. Granted, if one is to be a criminal master mind then one must display special care in executing ones schemes. One cannot maintain good standing qua criminal mastermind while feeling disinclined to display the degree of sophistication in plan and execution constitutive of it. But why be a criminal master mind in the first place? Why internalize the set of standards that characterize that mindset as opposed to some other mindset or no mindset altogether? Here we finally encounter the kind of

question more apt to our overall meta-normative inquiry. And it is exactly this kind of question that can be raised with regard to belief as well. The crucial question is not why care about truth when believing (which is defined as exclusively caring about truth), but why to believe in the first place. Why must one be a believer, as opposed to a celiver or shmbeliever, or no XX-liever altogether? *Truth is the aim of belief, alright, but why ought it be my aim*, one may ask. Yes, if one respects the belief-constitutive norms one qualifies as a believer. But why should one then go ahead and respect the norms and thereby become a believer? Once the question is cast in those terms, it is immediately evident that the constitutive account has little to say in response. The constitutive account does not tell us why to respect the norms constitutive of believing, but only that if one does respect them then one is a believer. It does not tell us why to form acceptances in the specific belief-way, but only what that way is. What we have granted all along is that with believing comes a package of norms where one cannot have one without the other. The question, however, is why one must take on the whole package in the first place, and to that question the belief-internal and constitutive norms turn startlingly silent. Whether one should have the kind of special regard for truth that is constitutive of belief is not answered by the recognition that believing as such is to have this kind of special regard for truth.

This concludes the argument. Notice two points, though. First, one may complain that we do not really have a choice in the matter of whether we want to be guided by representations of the world we live in, and that consequently the question of whether we want to be believers is largely rhetorical for us. Not quite. Even if we grant the necessity of forming representations of the world on our part, the necessity of forming beliefs does

not immediately follow. To be a believer is more than to be someone who is guided by representations of the world. The example of believers made this plain. To be a believer is to be guided by representations of the world of a *highly specific kind*, and to that there are ample alternatives. Many incorporate less stringent standards. Moreover, the equation of inescapability and authority is highly problematic, as I will argue in later sections. The second point is this. I certainly have no intention of denying that possessing the belief-characteristic capacity is hugely beneficial to us, that it is a capacity we cherish to have. More often than not it is a very good thing indeed to be a believer. But that recognition is neither here nor there as far as the constitutive account is concerned in particular and the wider issue of authority in general. My contention with authority is not a contention with the norms defining our epistemic practices. And whatever explains why we wish to exercise, promote or foster that capacity will have to involve factors beyond the constitutive analysis. Trying to gain some argumentative leverage with regards to the authority of the norms underwriting belief on such an instrumental basis would lead us into a different project altogether, one that in large parts was dealt with in the previous chapter.

4.5. Identifying the Constitutive Base for Action and Agency – An

Initial Dilemma

Let us return from the theoretical to the practical, from reasons and norms pertaining to belief to reasons and norms pertaining to action. The overarching objective, we need to remind ourselves, is to explain authoritative practical reasons. It was not

entirely implausible to suppose, as the previous section did, that there are norms constitutive of believing. The question we must now consider is whether there are practical norms to be found which pertain to action and agency and which could claim the kind of constitutive status for their respective domains that norms pertaining to belief were able to claim with some credibility. “The question, then,” Velleman notes, “is what serves as the standard of correctness for action, in the same way as truth serves as the standard of correctness for belief.” (2000:15) The task is simultaneously more challenging and more promising. More challenging, since one major quandary that agency-based accounts have always faced is to identify concrete norms that are plausible candidates for principles constitutive of agency and action. Rarely do we find concrete norms suggested in the literature that meet this job description, where major parts of the discussion proceed in entirely abstract terms, only alluding to norms that supposedly are constitutive of agency without saying what they are. Yet without this important detail it is hard to gain an adequate assessment of the merits of agency-based accounts, and thus at some point we need to know precisely what, according to agency-based accounts, one must honor in order to be an agent.

At the same time the task is more promising, in at least one respect, for the following reason. While it is hard to deny that many practices come with their internal and constitutive standards that characterize the practices as what they are, one question always relevant to ask is why engage in these practices so construed in the first place. With regards to belief, the question was why be a believer instead of a deliever or something else? Since most practices are escapable, their constitutive standards become escapable as well. Now we have not examined the exact relationship between authority

and escapability yet, which we will do in a later section. For now it is worth mentioning that there may be one exception to the escapability of practices together with their corresponding standards, and this exception is agency and action, or at least so it is claimed by many proponents of constitutive accounts. If that is so, this would give constitutive accounts of agency an extra card to play. Unlike in the belief case, where it was possible to ask *Truth is the aim of belief, alright, but why ought it be my aim?*, this kind of question arguably cannot arise in the agency-case, where it appears nonsensical to ask *XYZ is the aim of agency, alright, but why ought it be my aim?* As I said, that additional option will be the subject of a detailed examination later. Return now to the task of figuring out which norms could possibly be held constitutive of agency and action.

Even the little we know about the content of the constitutive norms already spells trouble. An initial dilemma threatens. Constitutive accounts aspire to explain reasons for action or authoritative practical norms by situating them in necessary features pertaining to our agency and capability to act. Yet these two components pull in different directions, namely the aspiration to explain authoritative reasons for action on the one side and the excavation of an explanatorily fertile stock of necessary standards on the other side. The candidate norms underwriting agency can be construed in more or less stringent terms, and neither option promises to accommodate quite what we want. The first horn of the dilemma is this. Suppose the candidate set contains some fairly robust and stringent norms. It would then stand a chance of underwriting most of what is usually considered legitimate demands on action, explaining the target domain of authoritative practical norms and reasons. The difficulty here is to substantiate the strong necessity claim

characteristic of constitution. In other words, if the candidate norms involve substantial and stringent requirements, it appears more probable that agents may opt out from those norms without compromising their status as agents.

This leads to the second horn. Based on the just presented difficulties we may conclude that we have little choice but to construe the candidate set of constitutive norms for agency less stringently. Suppose, then, that alternatively the candidate set only contains norms that absolutely cannot be opted out from without relinquishing our agency. The difficulty here is to substantiate the claim that constitution is what explains legitimacy in principles without at the same time forfeiting the capacity to accommodate a fair deal of the principles usually considered legitimate vis-à-vis action. In other words, if the candidate norms are necessary for all agents to be agents, it appears their content must be very weak indeed and in all likelihood can accommodate only very few principles widely held legitimate. Perhaps proponents of the constitutive account may moderate their ambition at this point, where they seek to sidestep the trend into some normative wasteland by claiming that constitution is only part of the story. As we shall see, however, that is no easy concession for agency-based accounts to make. The first horn threatens the necessity claim behind constitution. The second horn threatens the claim to explanatory adequacy behind constitution, allowing, at best, only the minutest set of principles the status of authority. Agency-based accounts face the uneasy task of capturing authority and necessity at the same time.

This is a genuine challenge. Ultimately agency-based accounts have no choice but to make do with the second option, to resort to a small number of basic constitutive

norms. I shall argue momentarily why this option is unattractive. Still, it is less detrimental to agency-based accounts than the first option which plainly appears logically ruled out. The reason why the first option is indeed ruled out becomes apparent once we apply a process of elimination. Upon closer examination virtually the entirety of principles considered legitimate must turn out optional when it comes to maintaining our agency. The endeavor of finding necessary norms underwriting agency is aggravated by the sheer abundance of agents and their tendencies to flout almost all imaginable norms and principles. It is easy to find counterexamples for most norms restricting their inclusion to the agency-constitutive set for most norms. There are lots of bad agents and there are lots of instances of bad behavior. And as a matter of logic alone no norm that is systematically disobeyed by some agent can serve as a candidate for underwriting agency as such. Henceforth a great number of norms considered valid cannot be encompassed within the constitutive set.

Moral norms are just the most obvious case. Many have pointed out that the constitutive programme proves incapable of accommodating moral norms. Ever since the publication of *Sources of Normativity* Christine Korsgaard has been confronted with the evil gang of mobsters, murderers incorporated, sociopaths and psychopaths, and so on, who are all agents, albeit of the most horrendous sort. Looking at only a relatively small number of moral monsters we can already appreciate how small the intersection must become of moral norms they all respect. Perhaps one may object that one cannot lead a fully satisfactory life while systematically disregarding others. I am genuinely unsure about this, given the apparent success of numerous ethical barbarians to compartmentalize their social life and to flourish. But the objection is beside the point

anyway. The question is not what is required for happy or flourishing agency, but only for agency per se. Discontented agents are still agents. Out the window seems to go moral reasons for action.

Christine Korsgaard disagrees. She has recently argued that Kant's categorical and hypothetical imperatives figure as the two constitutive principles for agency and action. I'll comment on the hypothetical imperative in the section on Korsgaard later, and focus on the categorical imperative now. Korsgaard contents that "[t]he categorical principle ... is not just the principle of morality. It is also the constitutive principle of action." (2008:12) The categorical imperative may seem quite demanding and to thereby secure a place for morality within the constitutive set. Yet her constitutive claim strikes me as wholly incredible, in particular her equation of moral norms and norms concerning self-constitution, as expressed by the astonishing claim that "[t]he unity that is essential to agency and moral integrity are one and the same thing." (2008:14) There are many questions about the categorical imperative, what it says, how it works, whether it strictly speaking can ever be violated, how the requisite practical contradiction is generated, why to place it at center stage in morality instead of other principles, consequentialist, or contractualist, or whatever. None of this needs to bother us now. What should, instead, very much bother us now, is this: How could anyone seriously doubt that whatever the categorical imperative commands, and to the extent to which its commands can be violated, they are violated left and right by people looking very much like agents? Take promissory norms, something Kant, as well as other Kantians following him, including Korsgaard, used as one prime examples for norms backed up by, or instantiating, or expressing, the categorical imperative. Yet agents break promises all the time, and by

extension disrespect the categorical imperative all the time. If the categorical imperative is constitutive of agency, denying claimants the status when defying the imperative, none of us are agents, which would be a rather dubious result for a theory of agency, as Korsgaard acknowledges: “it is much harder for skeptics ... to deny that agency exists.” [than that free will exists] (2008:11) I shall return to a more detailed assessment of Korsgaard’s views after I discuss Velleman’s, a task to which I now turn.

4.6. Constitutive Norms for Action: David Velleman’s Narrative

David Velleman has developed a highly original version of the constitutive programme in a series of articles and books published over several decades, from his *Practical Reflection* to his most recent monograph *How We get Along*. This section is dedicated to assessing that version. First I need to preface the ensuing discussion by a cautionary note, however, similar to the discussion of the aim of belief. It is not obvious, and in fact there is some cause for doubt, that the most charitable interpretation of Velleman’s work on agency and action incorporates the aspiration to explain authority. Recently I’ve had the opportunity to ask him about a complaint he is confronting with great regularity, namely that his account is void of direct normative implications and is purely motivational in character, and his response was rather telling. He seemed quite happy to fully concede the complaint. *Yes, it is all about motivation*, he said as far as I can recall, *but motivation is all there is anyway*. The remark confirmed a suspicion emerging while studying his work. In light of his response to my question I hesitate to construe my critical observations as directed against Velleman’s project as such. Still,

and as long as we keep in mind this interpretative proviso, what he says about agency is certainly capable of bearing a critical assessment within the present context, which is organized around the topic of whether the authority of practical norms can be explained on the basis of what is constitutive for agency and action.

Let me begin with an important passage in the Introduction to *The Possibility of Practical Reason* where Velleman highlights a fundamental problem (2000:15-16):

There is a temptation to think that the norm of correctness for actions is that they should be supported by the strongest reasons. But this thought leads into a vicious circle. What counts as a reason for acting depends on what justifies action; which depends on what counts as correctness for action; which cannot depend, in turn, on what counts as a reason.

The notion of norms for correctness for action, Velleman contends, is conceptually prior to the notion of a reason for action, and therefore is not analyzable in terms of practical reasons.⁶ This claim of his notably contrasts with the proposition, defended by several philosophers, assigning priority to reasons, both conceptually and explanatorily. Yet what Velleman attempts to do, again in stark contrast to his opponents, is to bring some illumination to the notion of a practical reason, instead of just taking it as a primitive. And once the nature of this commendable task is appreciated and accepted, we can immediately see why Velleman is right in searching for a criterion for correct action, and derivatively practical reasons, that must be found somewhere else than in the

⁶ Compare this with the case of belief, where Velleman's claim seems even more plausible: The norms for correctness for belief cannot just be to form only beliefs that are best supported by reasons for belief. What exactly provides reasons for belief and why is exactly what we would like to know.

mere generic notion of what is supported by reasons. Where to look, then, instead?

Velleman first gives a broad recipe before specifying the ingredients (2000:16):

If there were something at which action constitutively aimed, then there would be a norm of correctness internal to the nature of action. There would be something about behavior that constituted its correctness *as* an action, in the same way as the truth of a propositional attitude constitutes its correctness as a belief. This standard would not be open to question: actions meeting the standard would be correct on their own terms, so to speak, by virtue of their nature as actions, just as true beliefs are correct by virtue of their nature as beliefs. And this norm of correctness for action would in turn determine what counts as a reason for acting.

What, in particular, does Velleman propose as the constitutive aim figuring for action? The answer is somewhat surprising. The constitutive aim of action, so Velleman, is to know what one is doing, or to be intelligible (2000:26):

[S]elf-knowledge is the constitutive aim of action ...[which] determines an internal criterion of success for acting, in relation to which considerations qualify as reasons for acting. ... [T]he considerations that qualify as reasons for doing something are considerations in light of which, in doing it, the subject would know what he was doing. They are, more colloquially, considerations in light of which the action would make sense to the agent. ... When I speak of “making sense,” I am borrowing the phrase from the domain of theoretical reason, where it is used to characterize phenomena as susceptible to explanation and understanding. What makes sense to someone, theoretically speaking, is what he can explain. This is what I mean when I say that reasons for doing something are considerations in light of which it would make sense. I mean that they are considerations that would provide the subject with an explanatory grasp of the behavior for which they are reasons.

A decade later we still find the same central claim: “action consists in behavior that follows considerations that make it intelligible to the agent. Action is thus behavior aimed at intelligibility, just as belief is acceptance aimed at truth.” (2009:133) Explanatory grasp and intelligibility must not, of course, be understood in terms of what is most supported or explained by reasons, except in the stipulative sense that what provides explanatory illumination and intelligibility ipso facto qualifies as reasons. The terms in which the relevant explanation must proceed cannot be the terms of practical reasons.

Velleman’s proposal is subtle and part of larger philosophical projects, particularly within the philosophy of action. He is elaborating on some central themes of Elisabeth Anscombe’s, above all how we can acquire non-inferential and non-causal knowledge of what we are doing, as when one knows that one is going to take a walk just by forming the very intention to do so. Doing full justice to his proposal would require an adequate treatment of the kinds of problems in philosophy of action to which the proposal is responding, justice I must deny him here for reasons of space. The proposal has also prompted extensive commentary, mostly negative, where the common reaction, it is fair to observe, has been incredulity and the complaint of running against widespread intuitions. Two especially common lines of criticism, in Velleman’s own diagnosis, are that his view is “oddly intellectualist, or as portraying an autonomous agent to be unduly self-absorbed.” (2000:30) These are valid points of contention, which cannot be fully developed here for reasons of space, though I will briefly return to one of them in the concluding paragraphs of this section. The question I want to focus on, instead, is this: Is it plausible to suppose, with Velleman, that self-knowledge and intelligibility are the

constitutive aim of action? And, given that I'm inclined to respond in the negative, why not?

The question is somewhat speculative, as is to be expected given the nature of the proposal; still, there are some empirical matters that seem to bear on it. In a remarkable passage Velleman is quite frank about this (2000:17):

[T]he constitutive aim of action would have to be something at which it was in fact *aimed*; and its being aimed, in some direction or other, would be a fact about the mechanisms causing and controlling it – in particular, the mechanisms whose causing and controlling it were constitutive of its being action.

This is reminiscent of the purely regulatory account with regard to the aim of belief, and we already know that the proposal, as it stands and as it is cashed out entirely in terms of regulatory mechanisms, faces formidable obstacles in its attempt to capture aims-in-action. I shall not seize on that difficulty here, though, but instead focus on another that seems to arise even before we start worrying about the capacity of regulatory mechanisms to incorporate aims. The question is how plausible is Velleman's universal conjecture about the mechanisms governing action, as far as that goes, when confronted with the empirical data. We all have some intuitive, though defeasible, familiarity with action where we, at least occasionally, clearly know when someone acts, including ourselves. It would seem, then, that once we have identified a few uncontroversial instances of someone acting we could proceed by taking those instances apart and, by looking inside, we could check whether the relevant mechanisms do have the structure and functioning Velleman supposes. Now admittedly this is a much harder task than

suggested by the metaphors of taking apart and looking inside. Still, especially under the chosen mechanical description, Velleman's proposal clearly presents itself as a matter on which empirical disciplines such as psychology have something to say, and it is rather optimistic to forecast a confirmative verdict. Now Velleman does not shy away from facing up to empirical inquiries, and occasionally cites work in psychology in his favor. Providing a few cases in point he writes (2007:xvii-iii):

Strange as this psychological mechanism may sound, it has been copiously documented by social psychologists working in the area that is sometimes labeled self-consistency. Research in this area has shown that people have a broad tendency to behave in ways that cohere with their conception of themselves Potential voters are more likely to vote in an election if they have antecedently predicted that they are going to. Children are more likely to be tidy if told that they *are* tidy than if told that they ought to be. People behave angrily if they are led to believe that they are angry ... Shy people don't behave shyly if they are led to attribute the symptoms of their social anxiety to other causes. And so on.

I share Velleman's fascination for this kind of research and would readily grant him that there is indeed ample documentation that our self-conception, and thus in some way our concern for coming to understand what we are doing on our own terms, does often play a crucial role in the etiology of what we are doing. But as instructive as the results may be they can only be taken so far. Notice that what Velleman needs psychology to bear out is not just that we can find a drive for intelligibility underlying some substantial number of our motivations. What he needs psychology to bear out, rather, is that this would hold for all of our motivations, or at least for all motivations underwriting action as opposed to mere behavior or mere activity. For each and every

instance of someone acting we would need to find the corresponding aim in the corresponding mechanisms, and that, it is fair to say, would be a remarkable discovery. It's the kind of discovery that would warrant a Nobel prize, had only Nobel decided to award a prize in psychology.

Now there's a sense in which Velleman's empirical contention appears almost right, but that sense is rather unhelpful. Everything that happens has an explanation. If acquiring an understanding of why something happens or why someone acts suffices to render it intelligible, then for every action we can find considerations that render it intelligible: the considerations, namely, that are implicated in its etiology. Yet this, of course, cannot be what Velleman has in mind, since it turns the criterion for intelligibility into an unhelpful triviality, and, moreover, profoundly undercuts the appropriateness of any aim-lingo to action-guiding mechanisms. First of all, the fact that every action is explainable, perhaps even intelligible, cannot establish that every action aims at anything, let alone that every action constitutively aims at intelligibility. All of us are sure to have to adjust their ambitions in the course of our lives, but when ultimately forced to do so it is unlikely that we are thereby setting out to satisfy a goal of ours for self-adjustment. Moreover, the universal explicability of everything undermines any meaningful counterfactual tracking-account, where the co-variability with explicability for any mechanism would be converging on zero. It is clear, then, that Velleman must target at a more interesting sense of intelligibility, where what is intelligible is not just what is explicable.

Over the years Velleman has gravitated towards a narrative interpretation of intelligibility: “Reasons for acting are the elements of a possible storyline along which to make up what we are going to do.” (2000:28) Velleman elaborations of the notion of narrative intelligibility and its significance are undoubtedly quite exciting and original, yet also present a thorny interpretative task, where it is unclear whether we are always dealing with one and the same theory in Velleman’s disquisitions, confronting labels seemingly used interchangeably yet also seemingly denoting somewhat different things, such as *knowing-what-one-is-doing*, *doing-what-is-intelligible*, *doing-what-makes-sense*, etc. (Dancy 2004b) I shall not elaborate on Velleman’s narrative interpretation. What is clear, though, is that whatever sense of intelligibility he has in mind must exceed a purely explanatory sense. Yet by making the sense more stringent we are also dampening the likelihood that empirical studies would bear out that intelligibility in that sense underwrites every action-guiding mechanism. The more plausible hypothesis would be that sometimes considerations of intelligibility play a crucial role, and sometimes they do not, a hypothesis plainly incompatible with the strong necessity claim behind constitution.

Now, apart from such empirically based incredulity, there are reasons closer to home to remain somewhat skeptical towards Velleman’s constitutive theory, namely its notable failure to comport with our common self-understanding of what is involved when we act and when we decide how to act. Since Velleman places such a high premium on self-understanding, this is not a worry he could take lightly. Narrative Intelligibility, whatever its precise nature, seems such an unlikely candidate for a constitutive aim underlying all of our actions and in charge of, even if only implicitly, all of our decision-

making. I added the modification “implicitly,” for the manner in which intelligibility organizes deliberation certainly need not be conscious and in the open, with intelligibility the explicit goal-in-view. According to the most plausible version of intelligibility-based deliberation, a point Velleman emphasizes on several occasions, intelligibility figures in the background, much in the same way as, from what we learned in the previous chapter, desires figure in the background according to the most plausible version of desire-based reasons. Still, we can always momentarily push what is in the background into the foreground, and we should not expect to be surprised too greatly by that transition, if indeed what was supposed to figure in the background was there all along. After all, we are not talking here about deceptive ways in which factors exert influence from the unconscious *id*. As deliberators, or as philosophers seeking to better understand deliberation, we can always make explicit what was supposed to be the implicit guide, and if under such conditions we remain utterly incapable of recognizing the contested feature as *the* organizing factor, the odds are high that it was not really the organizing factor after all. All this is defeasible, of course, but the best way, perhaps even the only way, to determine what does structure our deliberation is to make deliberation more explicit and see whether we can recover the impact of the relevant contested features. I see no real alternative to engaging in this kind of deliberative phenomenology.

Consider two sources of deliberative discontent with Velleman’s claim that intelligibility is the constitutive aim of action, the first regarding determinacy and the second regarding force. My primary complaint with regard to determinacy is not that Velleman’s account would permit for significant indeterminacy in what reasons or well-conducted deliberation call for. That’s alright in my view. The problem, rather, is that his

account seems to give way too quickly too often, where we appear to retain the ability to keep deliberating even after we have noticed that all options remaining on the table do equally well in terms of intelligibility. Choice-situations coming with an array of equally intelligible options are not uncommon. Here's one. I might decide to stay home with my book, music and wine-and-cheese, or I might decide to meet a friend and go out for dinner and a movie. Both ways of spending the evening seem perfectly intelligible, neither having the edge intelligibility-wise. Intelligibility seems neutral between the options, and yet I am still thinking. In a telling passage Velleman writes (2007:xxx [p.30 of Introduction]): "The standard of intelligibility is a standard of coherence. It requires an agent to find that action which is explanatorily most coherent with his values, commitments, motives, habits, customs, practices, and personality." But, as is well known, coherence is a rather permissive criterion. Both practical options of staying-at-home-reading-and-listening-to-music and going-out-for-dinner-and-a-movie perfectly cohere with my values, commitments, motives, habits, customs, practices, and personality. Now it is true that after I have made a decision the winner shall edge ahead intelligibility-wise. But that generic truth is not very helpful to consider before a decision has been reached, as it applies to all options equally.

Now consider force. Suppose staying at home is in fact the more intelligible option. Were I to write an autobiography, I would take great care in emphasizing how hoity-toity a person I am, and literature, music and wine-and-cheese better fits that kind of life impression I would then wish to give. The dish and movie I am sure to choose are rather ordinary in contrast, bloody and unrefined in both instances. Now intuitions turn thin here, but I'm inclined to find the stipulated intelligibility-imbalance a rather

insignificant detail for my decision-making-process, which could hardly be so if indeed it was *the* constitutive aim. It seems to me it would not take much to overrule it, perhaps a mental depiction of a savory burger at my favorite steakhouse would do. It seems a perfectly credible self-description to say that I chose what, on balance, was slightly less intelligible over that which, on balance, held greater culinary promise. Moreover, the voluntary surrender of some degree of intelligibility on my part would not strike me as too bothersome. Intelligibility does not appear to have the deliberative significance we would expect should it be *the* constitutive aim of action.

Now, it may seem fair to reply: wait a moment, would not the prospect of a savory burger have implications for what qualifies as the more intelligible course of action? Would not, quite generally, that which emerges as the pivotal deliberative factors for choice also contribute to that choice's intelligibility? Whatever considerations tip the balance in well-conducted decision-making is ipso facto also a consideration in light of which the decision reached becomes more intelligible. This potentially seems to undercut the provision of any counterexample against Velleman's account, where reasons tilt one way and intelligibility the other. I'm not quite convinced by this, given the relative ease with which stories can be concocted about sadists who take great pleasure in their cruelty or megalomaniacs who ruthlessly strive for power, rendering their reason-contradicting behavior intelligible in light of their motives nonetheless. Marquis de Sade, the eponym of *sadism*, was, after all, quite a gifted narrator of his own libertine lifestyle, portraying his excesses as no less understandable than any other conduct. However, few would conclude on that basis that he usually did what he had most reasons to do. Or, to take another literary example, when Meursault shoots a random person in Camus' *L'Étranger*

because it was so hot, his murder might be considered intelligible *to some extent* given the ample documentation we have concerning how profoundly seemingly irrelevant factors, such as heat, can affect behavior, especially when agitated. And yet that heat, whatever else it did, hardly provided any reason for Meursault to murder, not even *a little bit* or *to some extent*. The lesson, in very abstract terms, is that understanding and approval are not the same phenomena, often coming apart, even if also frequently going together. It is unfortunately not true that *tout comprendre c'est tout pardonner*. I am thus not convinced that Velleman's account is indeed free of counterexamples. But even if we should grant him that such counterexamples are impossible, this would still not necessarily count as a victory for Velleman, for intelligibility could be systematically correlated with the factors doing the real work without doing the real work itself.

And this is indeed not implausible to suppose. The rejoinder to the reply, then, is that Velleman's intelligibility account gets things backwards, where the manner in which intelligibility affixes to reasoned choices does not help his case. Once I have determined that certain considerations settle a deliberative case I have also made a choice rendered intelligible by the very same considerations. But intelligibility itself seems secondary rather than primary here. If I decide to go for the burger and the movie, because of their primitive appeal, I have made an intelligible choice. But the intelligibility of that choice is not the primary concern, but rather seems entirely derivative and a function of the primary concerns that are in fact driving the decision. It's a supervenient feature with little or no independent force. My choice had something to do with my appetite for burgers and primitive entertainment. Because it had something to do with anything, or anything appealing, it also was rendered intelligible by that anything. Now we may ask:

which is the more appropriate description of my deliberative situation? That, first, I was moved by considerations of intelligibility, if only implicitly, and that it so happened that a burger and a movie would make for a particularly intelligible action-sequence? Or, alternatively, that I was moved by the burger and the movie and that it so happened, quasi as a side-product, that intelligibility was thereby ensured as well? The second seems the more apt characterization.

Thus, even if intelligibility turns out to be a supervening feature of well-conducted deliberation, where intelligibility is always co-present when we reach well-reasoned decisions, Velleman is still mistaken in claiming intelligibility to be the constitutive aim of our choices. “The goal of a more comprehensive knowledge of what we are doing therefore militates in favor of doing things that can be understood as motivated by our desires, expressive of our emotions, implementing our policies, manifesting our characters, and so on,” Velleman contends. (2007:xxvi) But no, it’s the other way around. We are, non-derivatively and not usually in the service of comprehensive knowledge, motivated by our desires, striving for ways of expressing our emotions, seeking to implement our policies and to manifest our characters, and it so happens that thereby we come to be intelligible and known to ourselves as well. Self-knowledge is a resultant feature of purposeful action instead of its driving force. Of course, none of this is to deny that occasionally a drive for comprehensive self-knowledge might take a more active role, especially within social context where there is ample pressure to justify oneself. There is in fact an intriguing conjecture that our occasional obsession to justify ourselves *to ourselves* emerges as an offline rehearsal to justify ourselves *to others*. However this may be, and however wide reaching a concern

for intelligibility may be, what I emphatically question is that this drive is *the* organizing motive behind action and deliberation as such, without which they would not even be what they are. I am willing to grant the existential, but not the universal.

There are plenty other potential worries one could raise about Velleman's proposal. For reasons of space I had to restrict myself to developing only one line of criticism. In closing I'd like to mention at least one additional worry that seems especially pressing, and which incidentally is not unrelated to the argument from above. Velleman's proposal is at risk of systematically mischaracterizing the grounds of a great number of practical reasons by situating them in our own agency as opposed to situating them in aspects external to it. This seems especially pertinent when it comes to moral or other-regarding considerations. When one helps another person, and does so for the sake of the other person, the reason that would morally justify such behavior appears to have a lot to do with the other person and the opportunity to alleviating her plight and very little to do with the actor and the opportunity to further his overall intelligibility. Reversely, when one fails to help, in a morally culpable fashion, the primary moral complaint is best not understood as taking issue with the forgone opportunity of making oneself better understood, but rather as targeting at the forgone opportunity of alleviating the plight of another person. In the final analysis, the pivotal consideration here seems to have its source in the situation of the patient and not in the (intellectual?) situation of the agent, and, in particular, does certainly not derive its primary moral force from considerations pertaining to what would make for the most sensible storyline of the agent. Of course, various philosophers have challenged the notion of radically other-regarding reasons, potentially disconnected from the motivational set of the relevant actors. But whatever

one makes of those challenges, it is hard to believe that one is entitled to disallow the very possibility of such reasons on conceptual grounds alone, by citing some alleged constitutive aim of action which, per definition, appears rather self-regarding through and through.

What my argument contests is that intelligibility is *the* constitutive aim of action and agency; it contests that action has *that particular* teleology. For all that has been said intelligibility might still be considered a central and explanatorily fertile component in practical deliberation. My sense is that the final picture would be somewhat messy. Some people seem more concerned with their overall intelligibility than others. Concern for intelligibility comes in varying degrees, where we find different agential styles of prioritizing different kinds of considerations for deliberation. Some people worry about intelligibility a great deal, explicitly or implicitly, others not so much but instead focus on other things. I find it hard to believe that people of the first group are categorically deliberatively superior to people of the second group. I find it hard to believe, in other words, that intelligibility does have the privileged status of being *the* constitutive aim of action.

4.7. Constitutive Norms for Agency: The Integrity of Christine

Korsgaard

Christine Korsgaard has developed another version of the constitutive programme, resembling Velleman's in some respects while departing in others, but in any case quite unique. The resemblance between the discussion of both authors in focus and

criticism is sufficiently great, however, to warrant greater brevity here. Korsgaard's thought on the subject of agency and normativity has evolved from her *Sources of Normativity* in 1996 to her *Self-Constitution* in 2009, as well as her collection of essays in *The Constitution of Agency* published a year before, even though the basic ingredients and the basic problematic as she conceives it – her famous normative question – remained pretty much the same. I shall mostly focus on her latest work. The titles of her two recent books are intentionally programmatic and supply the central concepts of Korsgaard's approach to normativity and agency. In her view, the self qua unified agent is constituted through action. This is the entry point into matters normative: "the source of normativity lies in the human project of self-constitution". (2009:4) The crucial step she takes from this is that normative assessment of action itself must proceed in relation to action's capacity to constitute unified agency. Action is geared towards agency-constitution, and to be fully judged in terms of it. "If we want to learn what it is that makes action right or wrong, we must start by asking what actions are, what their function is." (2009:8) And, in a stunning completion almost 20 pages later, Korsgaard writes (2009:25):

Action is self-constitution. And accordingly I am going to argue that what makes actions good or bad is how well they constitute you. ...[T]he principles of practical reason are principles by means of which we constitute ourselves as unified agents ... that explains their normativity. The principles of practical reason bind us because, having to act, we must constitute ourselves as unified agents.

Her central thesis, then, is that the criterion of correctness for action is self-constitution. To recapitulate her view in familiar language, Korsgaard contends action is

constitutively aimed at self-constitution just as Velleman contended action is constitutively aimed at self-understanding. We saw why Velleman is mistaken. We now need to see why Korsgaard is mistaken as well.

My strategy against Korsgaard resembles my strategy deployed against Velleman, insisting on the phenomenological inadequacy of the picture of practical reasoning that emerges from her constitutive program. I believe Korsgaard's own work contains important elements to articulate my criticism most forcefully, enabling me to level my criticism quasi from within. Reasons of space permit to press only this one single complaint. But the problems with her account are myriad, and let me at least mention what would have been my second choice. Korsgaard's entire programme is premised on the claim that there is a function actions as such have, but actions as such do not have an overarching function or purpose or telos or ergon. Each and every single action has a purpose, for sure, since actions are teleological or goal-directed in nature. But to conclude from this that there must be a single overarching function or purpose underwriting action as such is to commit the same sort of quantifier mistake that Anscombe so beautifully diagnosed in Aristotle: namely that, supposing that for each action there is a purpose, we are entitled to conclude that there is one purpose for each action. We are not entitled to take that step, however, as moving from $\forall \exists$ to $\exists \forall$ is a fallacy. Now there are other arguments in support of the claim that there is a function-of-action-as-such, and Korsgaard defends her single-purpose teleology extensively, heavily borrowing from Aristotle. But the project seems doomed from the very start, for we have little choice but to consider Aristotelian teleology as discredited. To me it sounds odd even *prima facie* to ask what action as such is there for. I know what this or that action is there for, to quench

my thirst or to express my gratitude or to live out my passion. But there is no common denominator or purpose for all of my actions, except perhaps the generic one to achieve whatever particular purpose my actions have, which merely restates that each and every action has a purpose. Suppressing that fundamental dissention, I now turn to my more pointed criticism of the thesis that action has the particular function of constituting ourselves as agents.

Conceptualizing the normative force of standards in terms of their *necessitation*, a term she borrows from Kant, and a force in our psychic lives she complains is underestimated and misplaced by her opponents, Korsgaard writes in a particularly dynamic passage (2009:7):

There is work and effort – a kind of struggle – involved in the moral life, and those who struggle successfully are the ones whom we call “rational” or “good.” But it is not the struggle *to be rational* or *to be good*. It is, instead, the ongoing struggle for integrity, the struggle for psychic unity, the struggle to be, in the face of psychic complexity, a single unified agent. Normative standards – as I am about to argue – are the principles by which we achieve the psychic unity that makes agency possible. The work of achieving psychic unity, the work that we experience as necessitation, is what I am going to call self-constitution.

It is not entirely clear what Korsgaard has in mind when she speaks of *necessitation* and her repeated allusion to *psychic unity* and the *struggle* to achieve it makes it seem as though she is more concerned about matters psychological than matters normative, let alone meta-normative. Korsgaard’s famous normative question frequently threatens to turn into a psychological question, as if it concerned our capacity for self-

management and impulse control. This is not an uncommon complaint about her work (Hussain & Shah 2006, Parfit 2006, FitzPatrick 2005), and not entirely unmerited either, but for now let us set it aside and read what she has to say as a contribution to our present meta-normative inquiry, an attempt to explain authority.

Now I believe there is an important insight in Korsgaard's portrayal of self-constitution. I share her skepticism about pre-constitutional agency, of little selves that stand behind all of our actions and deliberations, of little selves that are standing ready for extra attributes to be attached to them, like dust collectors. I thus agree with Korsgaard that "there is no *you* prior to your choices and actions, because your identity is in a quite literal way constituted by your choices and actions," (2009:19) where the negated priority is understood principally instead of singularly. Of course there *is* a prior you to particular choices of yours, such as eating serials for breakfast this morning, but that is not how to understand Korsgaard's contention. Rather, there is no prior self to choice and action as such, but rather it is through choice and action that the self is self-forming.

On this issue we agree, then, but notice that it really is a side-issue of sorts as our present question is concerned. That question is not primarily about agency and its formative relation to action, but about action and its normative relation to agency. The question is whether action has the particular teleology Korsgaard supposes, namely of constitutively aiming at self-constitution, of owing its pertinent standards of assessment to what contributes to self-constitution. To see the distinction I am getting at more clearly, consider a simple analogy. A community requires people, but that doesn't mean

that people are there in order for the community to exist. Likewise, that agency is in some way made up of choices and actions does not mean that the point and purpose of choices and actions are to make up an agent. It is not true that a concern for agency sets the standards for action and dictates what to do. It is true that one acts and *thereby becomes* an agent, but false, in contrast, that one acts *in order to become* an agent. One can take a left turn or a right turn, one can join the Free French or care for his mother, one can stay at home with a book or go for a movie; each time our choices will, to a greater or lesser extent, determine who we are as agents, literally *making* us into who we are. But that is just the inevitable effect of making choices – disregarding for the moment particularly self-destructive choices – and does not supply normative standards for what particular choices to make. Action as such no more aims at self-constitution than it aims at self-understanding, even though it is heavily implicated in both.

I believe this relates to a noteworthy tension in Korsgaard's own theory. The tension is generated by newer and older elements of her theory, first by her new contention that normativity has its source in agency-constitutive standards and her older contention, which she still endorses (2009:20-26), that practical reasons and standards flow from our contingently-chosen identities. She first claims that "*the* principles of practical reason serve to unify and constitute us as agents, and that is why they are normative"⁷ (2009:27, emphasis added), but then she also claims that "our conception of our practical identity govern our choices of actions" (2009:20), and these two claims are

⁷ I read this as a universal claim, *all* principles of practical reason or principles of practical reason *as such* ..., thus the added emphasis on *the*.

harder to reconcile than she appears to acknowledge. Identities have a tendency to develop a normative life of their own, independent and not easily be brought back to that of agency and the pressure towards unification. Her older views strike me as closer on target, but in any case, they don't quite fit with her newer views, thereby revealing a major difficulty in taking self-constitution to be the standard of correctness for action.

Let me explain. Korsgaard's final picture appears to be this. Begin with the necessity of being unified agents, of coming to value and respect one's own humanity and rationality. The relevant principles here are those in virtue of which we are able to constitute ourselves as unified agents, enabling us to successfully struggle for unity in complexity. That's the ground level. But there are many alternative ways of being unified agents, many ways of acquiring identities, as fathers or professors or citizens or whatever, and her view she carries over from *Sources* is that many of our reasons in life flow from the contingent identities of ours we choose and endorse. Further up, therefore, we encounter choices and principles that are not strictly speaking necessary and dictated by our agency as such. The principles we are dealing with there concern what to do as good fathers or professors or citizens, but since no one needs to be a father or professor or citizen in order to be an agent, the relevant norms apply only contingently. Had one chosen another identity other norms would apply. Yet whatever identity one acquires, the reasons that as a result flow from it appear to exhibit no particularly strong tie to a concern for self-unification, but rather are concerned with what good fathers or professors or citizens would do in certain circumstances. They seem to introduce an independent source of normativity, and the question is how to make good on her claim cited above that *the principles of practical reason serve to unify and constitute us as*

agents, and that is why they are normative. The only way to answer that question appears to be the attempt to relate the normative significance of our identities to the necessity of being self-unifying agents, perhaps by grounding that significance in agency in some way, such that the normative significance of our identities ultimately is owed to the necessity of being a unified agent. Korsgaard in fact anticipates that move in her previous writing, as a component she is independently attracted to.

Unfortunately there are serious complications for how to relate the normative significance of higher-level norms to ground-level norms pertaining to agency, even if it turned out that living by those higher-order norms is our way of becoming agents. This is most dramatically seen in cases of conflict, where the ground-level norms and the higher-level norms pull in different directions, undermining the hope that the force of the latter could be explained on the basis of the former. Return to the example of being a father. A father ought to dedicate significant portions of his time and energy towards furthering of his child's well-being. This is what the higher-level norms pertaining to fatherhood are calling for. However, doing this may very well conflict with the project of becoming more of a unified agent. Dedicating the required time may put strains on ones unity, and abandoning ones fatherly duties may render it easier to become more of a unified agent. It may or may not, it all depends on the situation, but in any case it would hardly surprise, since fatherly norms are tailored towards an object other than the father, telling him to do something about his child, while self-constitutive norms are tailored towards his self, telling him to do something about his unity and constitution. The crucial point is that living up to the higher-order norms that flow from ones identity need not necessarily further, and could even compromise, one's self-constitution. Dealing with life's

complexities while preserving ones unity in a manner that comports with ones identity is only one way of doing so. Another is to run away, or to switch identities, or to acquire a unified identity as a bad person. If so, how could the necessity of preserving ones unity call for honoring ones contingent identity? Reversely, and in particularly extreme cases, living up to ones identity and self-embraced standards might even foreseeably result in one's total self-*disintegration* and death, assuming here that self-*disintegration* is a particularly poor way of achieving self-integration. Sacrifices by parents on behalf of their children are not unheard of, and neither are people who give their lives to all sorts of causes, thereby giving up their own *selves* on behalf of something else in selfless fashion. Henceforth, what our contingent identities call for and what the drive for self-unification calls for can easily come apart. Since the necessity of being a unified agent does not directly translate into any corresponding necessity to be an agent of a particular identity also living up to the relevantly accompanying standards, and since the promotion of greater unity might even take a hit by the commitment to ones identity, the normative force of the latter cannot be comfortably derived from to that of the former.

What I conclude, then, is that the standards that seem most pertinent for action and deliberation are not first and foremost standards concerned with self-constitution and self-unification, and are neither derivative of them. As in my discussion of Velleman, I question the phenomenological adequacy of the picture of deliberation emerging from Korsgaard's constitutive program, trying here to pitch elements of Korsgaard's own work against herself. Though this line of criticism could easily be extend, adding example after example, I won't do so here, but instead isolate this problem and turn to another. Suppose I am correct that norms pertaining to self-constitution cannot underwrite norms for action

in its entirety. What about agency as opposed to action, quasi as an independent practical domain freed from the need of having immediate implications for action? Are there principles that could possibly be held necessary for the possibility of agency as such? In good Kantian fashion, Korsgaard proposes two principles, the categorical and hypothetical imperatives. I already had something to say about the capacity of the categorical imperative to be recruited to that end, where I found it utterly incredible to suppose that the categorical imperative is constitutive of unified agency as such. Why is it also implausible to suppose that the hypothetical imperative is constitutive of unified agency as such, even if, admittedly, much less implausible than in the categorical case?

Suppose the hypothetical imperative directs us to adopt necessary means to our considered ends. In the final analysis the imperative would have to be infinitely more complicated, but for now, this simple version shall suffice. What would a constitutive account of the hypothetical imperative look like? It would say something such as this: it is constitutive of willing that one seeks out means of accomplishing what one wills, and in case something turns out necessary for what one wills, the willer cannot disregard that thing altogether while still being said to will what he wills. With Kant one could hold that “Who wills the end, wills (so far as reason as decisive influence on his action) also the means which are indispensably necessary and in his power.” (*Groundwork*, Paton translation), and, momentarily disregarding Kant’s own qualification, read this in the following way: *Who wills the end, wills also the means which are indispensably necessary and in his power, or he does not really will the end after all*. Korsgaard paraphrases the same idea (1997:244, emphasis hers):

To will an end just is to will to cause or realize the end, hence to will to take the means to the end. This is the sense in which the principle is analytic. The instrumental principle is *constitutive* of an act of the will. If you do not follow it, you are not willing the end at all.

In another passage, she writes: “If acting is determining yourself to be a *cause* of some state of affairs, then you are *just not acting* unless you take the means to that state of affairs.” (2008:15) We might go further, and hold that the hypothetical imperative is not just constitutive of willing, or acting, but of being an effective willer or actor as such, of someone who is capable of forming and executing his wills. The thesis would be, then, that one cannot be an actor, one who executes what he wills, while disrespecting the hypothetical imperative. Korsgaard formulates the idea (2008:13):

The notion of efficacy brings in the other element of Kant’s account of action, the principle of instrumental reason. For if to act is to engage in practical activity that is directed to producing some state of affairs in the world, then the agent must also seek to be efficacious, that is, to work with the natural causal mechanisms that he can use to make things happen in the world. He must use the means. And this means that the maxim or principle on which he proposes to act must serve as a universal *practical* law. ... To be an agent is to be, at once, autonomous and efficacious – it is to have effects on the world that are determined by yourself. By following the categorical imperative we render ourselves autonomous and by following the principle of instrumental reason, we render ourselves efficacious. So by following these principles we constitute ourselves as agents: that is, we take control of our movements.

Unfortunately the constitutive reading of the hypothetical imperative won’t do, and Korsgaard is keenly aware of the primary reason why. Rendering the hypothetical

imperative constitutive of willing, or of being an actor, immediately renders instrumental irrationality impossible. But it is not impossible, since people often do behave instrumentally irrationally, and then we cite the principle to ground our criticism. Korgaard writes (1997:236, or 2008:48):

Kant says that imperatives are expressed by an *ought* because they are addressed to wills that are not necessarily determined by the objective laws of reason. ... In other words, imperatives are addressed to beings who may follow them or not. And this is true of the instrumental principle as well as of the others. Now if this is right, it must be possible for a rational being (one who is subject to the instrumental principle) to disobey, resist, or fail to follow that principle. It must be possible for someone to will an end, and yet to fail to will the means to that end.

In response to the problem, while still attempting to hold on to the constitutive reading, Korgaard proposes to weaken the constitutive proposal, where willing does not constitutively require the *successful execution of the requisite means*, but only *the formation of a first-personal commitment* to adopting the requisite means (1997:245):

[W]illing an end just is *committing* yourself to realizing the end. Willing an end, in other words, is an essentially first-personal and normative act. To will an end it to give oneself a law, hence, to govern oneself. The law is not the instrumental principle; it is some law of the form: Realize that end. That of course is equivalent to ‘Take the means to this end.’ So willing an end is equivalent to committing yourself, first-personally, to taking the means to that end.

The passage raises as many questions as it answers, and I’m not fully confident I’m getting the idea. But however we read the weakening of the condition from *actually adopting the means* to merely *committing oneself to adopting the means*, the proposal is

simultaneously too strong and too weak, a result we also get for many alternative ways of weakening the condition, for instance *merely trying* to adopting the means, or *merely seeking* to execute the means. The modified condition is too strong, because the modified proposal rules out the possibility of *committal* instrumental irrationality (and, mutis mutandis, irrationality with regards to trying and seeking), which is in fact quite possible, even actual. All we need in order to counter this modified version of the constitutive reading of instrumental rationality is to modify the criticism, an easy exercise. People can and do fail in forming the requisite commitments to their ends.

The proposal is also too weak, however, for one does not take care of everything one needs to take care of in order to meet the hypothetical imperative by forming the suitable commitments. It is not enough to say: 'I am willing the end and I am also committed to executing the means,' and then stopping short of actually executing the means. To the extent to which *being committed to executing the means* is different from just *executing the means*, what one needs to do according to the principle is not just *being committed to executing the means*, but actually *executing the means*, and one fails with regard to the principle if one merely commits instead of executes. Perhaps one is inclined to reply that one cannot be committed to executing the means while stopping short of executing them, for not stopping short is constitutive of the commitment. Not so, as by now we know, and need not repeat why.

Neither the categorical nor the hypothetical imperatives are plausible candidates for principles constitutive of agency, let alone action. Still short of constitutive principles for action and agency we might start becoming discouraged. Is there not a single

candidate that stands a chance of constitutively underwriting agency? I think there might be one, and to this I now turn.

4.8. The Past and the Future, Michael Bratman's Plans and the Importance of George Sher's Past

Agency is a temporally extended phenomenon. Having a past and having a future are essential to agency, in a particularly profound sense of the term. It not merely so happen that agents traverse through time and thereby inadvertently become part of some ongoing history. A lot of things can be said to have a history, buildings and books, for instance. They persist through time and thereby come to acquire various temporal characteristics. Things happen to them that shape the way they are. Many aspects of us have histories that resemble the manner in which buildings and books have histories, our joints and bodies for instance. It doesn't even matter whether it was us who caused our joints and bodies to undergo certain modifications. The shape our joints and bodies are currently in reflects what we have done with them in the recent past. But that sense of acquiring a history is still not the interesting agential sense, as it is still exhausted by what happened to our joints and bodies.

The crucial further agential sense of having a history, then, must be going beyond the passive sense from above where things merely happen to agents, and where instead agents act and are not merely acted upon. The crucial entry point is this. Agents have a conception of who they are and under which they do what they do, representing a truly active sense of authorship of their own history. Agency cannot be time-sliced but

temporally extended and must be active. What we must above all avoid, however, is trying to capture the present sense of authorship by postulating little authors in our heads writing our stories, be it in the form of noumenal selves outside the causal order or be it in the form of some privileged subset of our attitudes, possible higher-order attitudes. Such a homuncular approach would offer no illumination for anything.

Active authorship over time must be understood differently, therefore, and the only sense I can think of that avoids postulating homunculi while also preserving activity instead of passivity is a sense of self-organization across time, underwritten by certain principles, instantiating a distinct form of capacity. And what an astonishing capacity that is. Squirrels hiding their nuts and bears feasting for hibernation display a first instinctive approximation of that capacity. Yet their preparations for wintertime fade by comparison when we consider activities such as building a cottage and cultivating the land. For this a myriad of sub-activities need to be coordinated, always keeping the final product in mind, being prepared to making necessary adjustments without giving up on the overall scheme, forgoing temptations arising for the present self on behalf of the later self, and so on. This requires enormous intelligence, and the relevant thought is that the principles empowering that intelligence are, in a manner of speaking and at some fundamental level, not first and foremost principles *we employ* but principles *we are*. *Extended active agency* is but a name for the capacity of operating under the relevant principles. Notice that this involves no bootstrapping paradox where agents self-create. Instead, to be an agent, in this view, is simply to be a system that operates under the requisite self-organizing principles, a process where the past and the future become integrated into a contiguous unity.

This is reminiscent of much of what Korsgaard and Velleman have to say, and indeed, despite all the points of disagreement, the core assumption I share with proponents of agency-based accounts is one that rejects as inadequate the notion of agency as an antecedent unity awaiting further cross-temporal-integration, and which holds instead that integration and the principles constitutive of it are what create that unity in the first place. It is not that there are agents who also happen to possess cross temporal characteristics in the manner books and bodies do. They exhibit the constitutive capacity to identify with future and past stages of *that person* and thereby establish their own agency. Projecting agency over time and creating agency over time seem two sides of the same coin.

We get a very immediate sense of how essential that cross-temporal identification is when we envision the terrifying scenario where we lose our memories and our ideas of what to do next. Our agency would shrink to an instant point in time and come dangerously close to extinction. What might preserve our agency at such an instant may be our awareness that we must have had a past even though we no longer remember it, and that we must go on making choices even though we have no clue as to which choices to make. But the lost sense of agency is very palpable indeed.

Michael Bratman's planning theory of intention is of great help for appreciating the intriguing nature of the principles underwriting cross-temporal agency. Alluding to a trio of core features of human agency – reflectiveness, planfulness, and having a conception of one's agency as temporally extended, Michael Bratman writes (2000:35):

We are reflective about our motivation. We form prior plans and policies that organize our activity over time. And we see ourselves as agents who persist over time and who begin, develop, and then complete temporally extended activities and projects. Any reasonably complete theory of human action will need in some way to advert to this trio of features.

Bratman believes, correctly in my view, that especially the last member of the trio – beginning, developing, and then completing temporally extended activities and projects – is key to understanding human agency, its cross-temporal identity, and in particular to understanding the other two members of the trio, reflectivity and planfulness (2000:57):

The primary connection between a self-governing policy and the agent's identity over time is grounded ... in the policy's characteristic role in coordinating and organizing the agent's temporally extended life in ways that constitute and support Lockean continuities and connections. In this way, on my view, a broadly Lockean approach to personal identity over time can help clarify the nature of the agent's reflective endorsement.

Lockean continuities and connections, or Parfit's overlapping strands of psychological connectedness, are established between different stages of a person in virtue of, in backward-looking manner, memories, retained commitments and desires, and, in forward-looking manner, intentions, plans, and policies. With regard to the details of all the memories, commitments, desires, intentions, plans, and policies involved, there is ample space for contingencies and differentiation. Incorporating some mix of them is required for an identity to form over time, but there is no necessity for any particular mix. I believe, however, we can push all this one step further, and focus on one especially fundamental capacity involved in forming and executing plans over time as such, namely

what Bratman has called a conservative bias in diachronic reasoning and planning, in the hope of finding something *constitutive*.

Forming an intention today to do something tomorrow requires some confidence on my part that when tomorrow arrives I shall still feel inclined to do what I intended yesterday, simply because I so intended. For one thing, it would make no sense to take the first sub-steps today if I could not rest assured that I will follow through tomorrow. The even more fundamental point, however, concerns not mere effectiveness, as important as it is, but rather the very sensibility of forming intentions in the first place, which crucially depends upon my anticipated preparedness to follow through past intentions in the future; which depends, in other words, on my anticipated preparedness to feel rational pressure in the future towards taking seriously, in a committal fashion, intentions and policies formed in the past. Without that assurance I cannot even sensibly begin to form intentions in the first place. My present formation of a forward-looking intention is possible only under the anticipation of a backward-looking future self, and in particular, looking back not in a mere act of curiosity but in an attitude of upheld commitment.

In a remarkably insightful chapter, *Why the Past Matters*, belonging to a monograph on desert, George Sher offers a more rigorous development and defense of the idea. Sher's overall objective is to come to a better understanding of desert and its significance, and, given desert's essentially backward-looking nature, the chapter's contribution to the monograph is to answer the question of "how what is behind us can affect the normative status of what is ahead ... how future occurrences of no intrinsic

importance can acquire value from, or be made obligatory by, their relations to events that have already faded from the scene.” (1987:175-6) Sher proposes to adopt a deliberative perspective to answer the question. “To show ... that reasons retain their force in retrospect ..., I shall seek this view’s rationale in the demands of deliberation and practical reasoning.” And, outlining the general form of his argument, he continues (1987:178-9):

More specifically, I shall argue that if we denied that reasons retain their force in retrospect, we could not coherently form certain very familiar sorts of intentions. Since an ability to form these intentions is central to deliberation itself, this would mean that we could not coherently deliberate. But ... deliberation is unavoidable, so the view that reasons are retroactively efficacious must be unavoidable too.

Sher’s emphasis on what an agent must take for granted so that he can sensibly form intentions, in other words states planning for what cannot immediately be achieved but what instead must be carried out later, repeats the important insight from above. “Because an intention fixes one’s readiness to perform the chosen act, it is a kind of bridge between one’s present and future self. In forming it, the agent necessarily envisions his current readiness to act as carrying forward until the appropriate future moment.” (1987:180) The agent cannot endorse, and in fact must reject, what Sher calls principle (NA), *a feature of an act or event is never a reason for anything once the act or event is past.* (1987:182)⁸ The beauty of Sher’s reasoning is that the significance he claims agents must accord to the past is explained on the basis of the fact that agents

⁸ I read *reason* here as a *deliberatively relevant factor*.

cannot otherwise make deliberative provisions for the future. The agential necessity of expanding into the future entails a similar necessity of continuing the past. Because tomorrow I must take seriously what I have decided today, and be justifiably expecting as much, lest my present decision loses all point and purpose, I cannot be deliberatively future-oriented without also being deliberatively past-oriented. The reasoning is particularly effective since it is addressed to those who take no qualms with the significance of the future for deliberation but who take qualms with the corresponding significance of the past, by showing them that they cannot have the one without the other.

Not surprisingly, this argument involves various complications and invites further attempts to clarify it. The larger context would have to be considered, the defeasibility of the force of past intentions admitted and balanced with other considerations, and so on, all of which could easily embroil us in a lengthy digression. Not now. However complicated the final account may be, however many subtleties need to be considered, there must be some norm we operate under which demands, in rough terms, that we take seriously, in a committal fashion, what we have done and devised in the past, even though the past is over, as one recent United States' President has put it. The practical question of why take seriously, in committal fashion, what one has done and devised in the past cannot be anything but rhetorical for us qua deliberating agents. It is not that we are agents, and then wonder about our fundamental attitudes towards our past and future. Having the relevant attitudes is what endows us with the amazing capacity called *agency* in the first place. It appears, then, that we have finally found, even if only in rough outline, a principle constitutive of agency as such.

4.9. The Retreat

It's time to pause and to take stock. The previous sections surveyed which principles could possibly be held constitutive of action and agency. Though we did not end up empty-handed, uncovering at least one plausible candidate concerning cross-temporal agency, the overall result was rather meager, especially with regard to constitutive principles for action. It is clear, then, that agency-based accounts do not succeed in coming up with a rich enough set of constitutive principles that simultaneously captures the kind of agential inescapability distinctive of constitution yet also accommodates a reasonable portion of norms widely considered authoritative. Logically speaking, agency-based accounts are free to embrace the minimalist result, and to refuse to withstand the tendency into a complete normative wasteland where only the minutes set of demands retains a claim to validity. There is not a single author, however, that I know of who would be willing to take that route, and so we can safely ignore it. The conclusion we must draw, then, is that the agency-based programme falls short of satisfying both its initial constitutive and explanatory aspiration.

Perhaps this was to be expected. Perhaps only the most basic set of principles of practical reason ever stood a chance of maintaining a claim to be constitutive of agency. Agents are known to exhibit the most egregious failures of practical reason by refusing to take into account their own immediate future, their own deliberative health, their acknowledged necessary means to execute larger plans and projects, and so on. In response, proponents of agency-based accounts may acknowledge that their proposal was only intended to be part of the meta-normative story. Perhaps the emerging picture is this.

Some norms acquire authoritative standing due to their centrality in agency, and some other norms acquire that same standing due to other factors. The demand to be responsive to intentions formed in the past might be an example of the first kind, whereas not to torture babies for fun might be an example of the second kind. According to this picture, what is constitutive of agency does not comprise the entirety of all authoritative norms. It simply comprises a particularly interesting subset and explains why all the members of that subset have that meta-normative status. One may go one step further and argue that constitutive norms are particularly interesting because they hold authority for all agents. They are categorical, perhaps uniquely so. Other norms can be legitimately applied to certain agents only. Role duties come to mind, such as those affixed with fathers and professors. In contrast, it may be argued that norms which are constitutive of agency retain normative force no matter what, independently of any further particulars of their target domain, since no agent could ever escape them. Being concerned about the view that “the justification of action must either lead to an infinite regress of justifications or stop short at a contingent aim that remains unjustified,” Velleman writes in response (2009:126-7):

The way to avoid it [namely the problem just mentioned] would be to show that actions can be justified as such, in the manner of beliefs – justified independently of any contingent aim at which they might be directed. In order for actions to be justifiable as such, however, they would need a criterion of aptness or correctness that isn’t contingent, a criterion that applies to them simply in virtue of their being actions, just as the criterion of truth applies to beliefs simply in virtue of their being beliefs. So we have to look for a criterion of aptness or correctness embedded in the very nature of action.

What this picture suggests is that constitution accounts not for a necessary condition for authority, but for a particularly interesting and powerful sufficient condition, since it also explains categoricity.

Given where we are in the overall dialectic, this strikes me the most charitable interpretation left. Nonetheless, the proposal comes with considerable costs that arise from within the agency-based programme. Recall how agency-based accounts were introduced and motivated in the first place, where the normative question was interpreted as a quintessentially agential question, and recall in particular Connie Rosati's ingenious interpretation of the normative problematic that put it in relation with Moore's open question argument. Her suggestion was that many accounts of what we must do and what is good for us retain an open feel precisely because they circumvent our own agency. Rosati complained that in particular many traditionally naturalistic accounts, such as naturalistic hedonism, involve principles that appear superimposed from the outside since they can be put at a distance from our own agency. What Rosati's line of reasoning suggests is that it is a necessary condition for a principle to retain authoritative force that it does not circumvent our agency. And according to the initial understanding of agency-based accounts they nicely captured this condition for it is not even possible that constitutive principles circumvent our agency for they enable it in the first place. Yet if now proponents of agency-based accounts are forced to amend their accounts by limiting their explanatory aspiration, permitting non-constitutive norms to retain normative force as well, the inevitable result is that we create a tension for this neat motivation of what the constitutive programme was all about. Indeed, once norms are permitted to retain authoritative standing vis-à-vis agents without a central place in agency as such, one

starts to wonder what the fuss about constitution was all about. If norms can retain normative force on agents, even categorical normative force – e.g. the categorical imperative – without being constitutive of their agency, then there are alternative ways to answer the normative question, and the condition of not circumventing our agency becomes rather optional, which is a polite way of saying that it is no real condition at all.

Thus, the original motivation, hooking up authority with agency-constitutive norms, cannot be so easily jettisoned without also jeopardizing the entire programme itself. And indeed, with the inclusion of non-agency-based norms we are forced to revisit the rationale for that very motivation itself, namely that a norm must connect to what is constitutive of agency to retain the right standing vis-à-vis deliberative agents. Return to Rosati who wrote (2003:507): “The kind of account of personal good that would escape such questioning [whether something is really good for us] would be one that could survive our reflective scrutiny, that would fit with our capacity for autonomous evaluation and action.” And also recall Korsgaard who wrote (1996:49): “If the problem is that morality might not survive reflection, then the solution is that it might.” The question we now face is this. Why must the process whereby we determine whether a norm survives reflection or reflective scrutiny, the process determining the norm’s standing vis-à-vis the agent, terminate in what norms we constitutively *must* accept instead of terminate in norms we *do* accept? Why is it not enough for a norm to fit our capacity for autonomous evaluation and action that it does have a grip on the actual agent? There is an imminent threat that the entire motivation behind the constitutive programme ultimately collapses into nothing but another version of the internalism requirement, in the process giving up on what is constitutive of agency and instead

retreating to the weaker condition that demands must find resonance in the contingent motivational structures and identities of agents in order to validly apply to them, the requirement, namely, that what demands must not circumvent is not some mysterious set of constitutive norms for agency, but rather that what they must not circumvent is whatever norms agents happen to reflectively endorse. A result Velleman in particular is at great pains to avoid.

This creates a problem so severe, I believe, that it warrants great pessimism about the internal stability of the agency-based programme after its explanatory ambitions have been modified – to which there is no real alternative. I shall not press this problem further, though, since there is an even graver one, and that concerns the move from the inescapability of certain norms for our agency to the authority of these norms. The move, crucial to this final and modified stand of the programme, emerges as fallacious upon scrutiny. The argument to this effect concludes the chapter.

4.10. Inescapability and Authority

What is the meta-normative significance of inescapability? The key question we need to ask is what conclusions we are entitled to draw about the meta-normative status of agency-constitutive norms based on the assumption that the sort of agency they enable is inescapable. Recall our present dialectical situation. We begin with the question whether norms constitutive of agency are authoritative and objectively reason-providing, and after being told that their authoritative standing traces back to agency, we are led to the question whether agency is authoritative and objectively reason-providing in the first

place. The operative principle is that norms can acquire authoritative standing due their role for agency only in case agency has the corresponding authoritative standing in the first place. For agency to lend authority it must already have it. And thus we ask: what is the meta-normative status of agency? In response to our question we are told that agency is inescapable. The claim is that agency, and derivatively its constitutive norms, are authoritative in virtue of being inescapable. The claim, however, is false. Agency's authoritative status cannot be established by showing it to be inescapable. If that assessment is correct, the agency-based programme is unsustainable and doomed in its very core.

Let's reconstruct the dialectic step by step, letting the programme's own proponents speak, beginning with the claim that agency and action is necessary. "Human beings are condemned to choice and action", writes Korsgaard (2009:1), and just as lively continues:

Maybe you think you can avoid it, by resolutely standing still, refusing to act, refusing to move. But it's no use, for that will be something you have chosen to do, and then you will have acted after all. Choosing not to act makes not acting a kind of action, makes it something that you do. This is not to say that you cannot fail to act. Of course you can. You can fall asleep at the wheel, you can faint away, ..., [a]nd then you will fail to act. But you can't undertake to be in those conditions – if you did, you'd be faking, and what's more, you'd be acting, in a wonderfully double sense of that word. ... [A]ction is necessary.

Velleman, reasoning similarly, takes this one step further (2004:290-1):

There may be, in some sense, an open question whether to be an agent, whether to get into or stay in the agency game. But of course someone

who is not already in the game is in no position to entertain that question, because entertaining it entails thinking about what to do, which entails trying to bring his behavior under descriptions that would embody knowledge of what he was doing. Anyone who asks himself whether to get into the agency game is already in the game; and anyone who asks himself whether to stay in the game cannot answer in the negative without staying in it at least that far.

Action and agency are supposed to be inescapable in the sense that we cannot get out from within. *Thrown into the role of agents* we are *condemned* to remain agents. Agency is not inescapable as a metaphysical condition, but rather agency is inescapable for agents, where no agential exit strategy is in sight. What does this kind of inescapability entail? David Velleman gives this answer (2009:116):

The view contends that such practical thoughts can be objectively valid without having indicative proxies that are made true by swatches of the world: practical thoughts can be objectively valid by being inescapable, in the sense that any agent must accept them.

Lucca Ferrero concurs (2009:304):

[The] basic claim is that the norms and requirements of practical rationality and morality can be derived from the constitutive features of agency. Hence, a systematic failure to be guided by these requirements amounts to a loss of agency. But there is a sense in which we *cannot but be* agents. It follows that we are necessarily bound by the oughts of rationality and morality, we are bound by them *sans phrase*.

What we have thus arrived at is the central claim behind the programme that norms constitutive of agency are authoritative because they are inescapable. We learn that such-and-such norms are constitutive of agency, that we cannot be but agents, and

that consequently we cannot but honor the relevant norms, and so we are invited to move from necessity to authority.

Yet what kind of necessity is in play here? Korsgaard starts out raising the right question. *Action is necessary*, she writes, and then proceeds to ask: *What kind of necessity is this?* (2009:1) After ruling out logical, causal and rational necessity, she concludes: “It is our plight: the simple inexorable fact of the human condition.” (2009:2) Agency, she believes, is necessary in the sense that it is the plight of the human condition, and for the sake of the argument I shall grant this to be so even though the precise contours of that sense are somewhat enigmatic. What is open to question, however, is the normative and meta-normative relevance of this kind of necessity.

Language seems to be coming to Korsgaard’s aid, making the transition from necessity to authority appear smoother than it actually is, owing to the fact that the meta-normative status of authority seems easily translatable into a kind of necessity too. After all, the normative question of what *ought I to do?* seems similar to the question of *what must I do?* This is even more salient when we use instead the all-purpose modality of *having-to-do*. Is not what *I ought to do* that which *I have to do*? And, reversely, if you have no choice but doing something, with no available alternatives in sight, then you *have to do it*, right? But then, is it not also the case that you *ought to do it*? To my ear, this already rings false, causing a first feeling of uneasiness in trusting language as a guide. The suspicion is that the modifier *have-to-do* is susceptible to normative and non-normative interpretations, and that moving from one to the other is a form of equivocation. We need to watch out for seductive yet illicit linguistic maneuvers.

But there's another intuitive route to Korrsgard's conclusion. Aren't normative questions practical questions of what to do? Isn't normative reasoning a process whereby we eliminate available options for choice until we have narrowed it down to a single choice which we thereby identify as the one we ought to choose? If so, then, in case there is only one choice to begin with – the inescapable choice – should not that choice also be the one we ought to do? If I wonder whether I should go through that door, and then learn it is the only door to go through, the question of what door to go through has been settled, hasn't it? It is tempting, then, to regard the step from necessity to authority an innocent one to take.

Yet it is anything but. To see why, start with a toy example, rolling a dice. Suppose there is the feature of the good or right number. You don't know which, if any, number has that feature. Now you roll the dice, and get a three. Have you rolled the good or right number? You don't know, but upon inspection find all sides of the dice are threes. You could not have but rolled a three. Does that answer your question? The answer is no. It tells you which number you had to roll, but why suppose that any number had the feature of being the good or right number in the first place? Now if you knew that, first, there was indeed a number that was the good or right number, and that second, it must be possible for you to roll that number, then you would be entitled to conclude you rolled the good or right number. But you don't know whether there was a number that was the good or right number, and you cannot infer this on the basis of what was inevitable.

Compare this with morality. It is intuitive to suppose that occasionally we encounter a range of morally neutral options. Driving home, I might take the short route or I might take the scenic route, and there's nothing more to it. There is no moral presumption in favor of either, let alone an obligatory presumption. Morality is simply silent on my choice of which route to take. Now this won't change if, for some reason or other, my choice situation is narrowed down to one option, due to road maintenance preventing me from taking the short route, say. If there was no moral presumption in favor of either option before, there will be no moral presumption for the single option now rendered inevitable. It is true that I have to take the short route. But it is not true that I am under any obligation to take the short route. Now we need to hear this right, since saying the latter usually involves all sorts of conventional implicatures, suggesting there must be other options I may permissibly adopt. But those implicatures are easily cancelable. In our example, the notion of moral obligation is not contrastively used, but simply presumed inapplicable. The important point is that the notion of obligation does not suddenly interject itself once we limit the option-set to a single choice.

Compare this with norms. Suppose there is the question of whether any norms in objectively authoritative. You then learn that there is a norm you cannot but implement. Have you thereby learned that this norm is objectively authoritative? Once more, the answer is negative. You would if you were entitled to suppose that, first, there indeed was a norm that was objectively authoritative, and that, second, it must be possible for you to implement that norm. But you don't know whether there is a norm that is objectively authoritative, and again you cannot infer this on the basis of what is inevitable.

There are various principles in connection with fairness and responsibility, which may seem to apply here and which may seem to support the notion that the inevitable choice must be the right or good choice. The principle '*ought implies can*' comes to mind. Now there are many complications with that principle, and in particular, there are complications concerning the relationship between evaluations we wish to make of choices and evaluations we wish to make of agents performing those choices, and the relevant principles seem most pertinent with regard to evaluations of agents performing the choices rather than of the choices themselves. But this is a hugely complicated and contested matter, and so I better stay out of it here; which I can comfortably do anyway since even without drawing the agent-choice distinction, the relevant principles cannot be recruited to support the contested inference from the inevitable to the obligatory.

Consider '*ought implies can*.' This principle does not have the desired implications of showing the inevitable to be the obligatory. All the principle says is that if there is an option someone ought to choose then that option must be one he can choose. And, under the assumption that there *is* an option he ought to choose, and further that there only is one option he can choose, this must indeed be the option he ought to choose. But the principle is silent on whether there always is, or even ever is, an option we ought to choose. The principle kicks in only after an option has been identified as the one he ought to choose and the principle cannot establish that there is such an option he ought to choose in the first place. The principle says something about all the oughts there are but doesn't say anything about whether there are any oughts. The principle '*ought implies can*' does not even show that there always must be one permissible choice. It shows something significantly weaker, namely that not all of the available options can be

impermissible. Permissible and impermissible are logically related to each other not as contradictories, where everything is either one or the other and never both, but as contraries, where nothing is both, but where some things might be neither. '*Ought implies can*' has implications for what is permissible and what is impermissible once permissible and impermissible are in play. It cannot introduce them into play itself.

The previous examples proceeded on the basis of inevitable yet normatively neutral choice-situations. Can we go one step further? I believe we can, and conceive of inevitable and normatively deficient choice-situations. Original sin comes to mind, possibly a morally suspicious concept, but surely not an incoherent one in its presumption that we all are inevitable sinners who cannot help but sin. The inevitability of sin does not affect its normative valence, nor does it provide any absolution. Catholics are not alone in this assessment. Especially pessimistic Calvinists might go further and consider human agency inherently corrupt and evil, a verdict they surely would not be prepared to revoke should agency be proven inescapable. More sanguine minded Protestants might conclude that we cannot blame and hold responsible those who cannot help but sin, but even they would not deny that what these people are doing is sinning nonetheless. From bad turn to worse, and consider Satan. Suppose it was him who engineered human agency, and, taking no chances in making sure that humans always sin, he resolved to build a drive for sinfulness right into the very design of agency, making sinfulness the constitutive aim of action and agency. How about that for a condemnation! Obviously no trace of normative worth and merit is contained.

Return from theology to agency, necessity, and constitution. The kind of necessity that underwrites constitution, whatever it is, is not the same that underwrites normative authority. The question is why we *should* be agents, what normative significance agency has, and the answer that we *must* be agents is neither here nor there as that question is concerned. It is not an answer of the right form to speak to that question. The space of modality *should* and *authority* are part of is not the same space of modality *must* and *inescapability* are part of, to which the fallacious nature of the inference *I must do X, X is inescapable* hence *I should do X, X is authoritative* bears immediate witness. Alternatively put, *should* and *authority* essentially are normatively commendatory whereas *must* and *inescapability* are normatively neutral, and no normative distinction may be inferred on the basis of what can and cannot happen, including what can and cannot be done by us.

Constitution concerns what we can't help doing more than what we should be doing. It belongs more to the purview of the engineer who designs complex norm-consuming systems than to that of the ethicist and practical philosopher. Suppose acceptance of norm N turns out to be constitutive for doing X. In that case you cannot do X while disrespecting N. Should you also care about doing X, then you would only achieve what you care about so long as you implement N (assuming that you did not have reasons to care). *You gotta do what you gotta do*. Yet nothing more seems to follow. In particular, nothing of the kind follows that you normatively must comply with N or that you have any reasons to do so. The impossibility of *doing-X-while-disregarding-N* could reflect some basic constraints in design-space akin to that one cannot build stable bridges while disrespecting the laws of gravity. Moreover, the inescapability of doing X itself

would not change the normative situation either, but solely place yet another constraint on your practical options. Thus, even if you absolutely had to do X – a choice you simply could not evade – and further that doing X required complying with N, this hardly would entail any reasons to comply with N on your part. One constraint would lead to another, but the fact remains that being constrained is not the same as having reasons. You merely would find yourself trapped in a tight corner. Might does not make right, however, as normative force evidently differs from brute force. Demands do not gain authoritative standing solely in virtue of their possession of absolute powers over us.

To conclude, inescapability and authority are very distinct properties. Once the distinction is appreciated, it should be clear what conclusion to draw about the agency-based programme in its own best case scenario; that is to say what conclusions to draw should all the myriad “smaller” issues and concerns threatening its viability be satisfactorily addressed. What we should say is this: there is a set of unique norms, special and interesting in many ways, that we agents cannot escape while retaining our agency. These norms constitute who we are, and we must, in virtue of some unique form of necessity, be who we are, namely agents. This is how we are built. This is how nature has shaped up. If you try to shake off your agency you will fail. What about the normative and meta-normative question concerning their authority? That question is entirely left open. All three propositions are consistent: The agency-constitutive norms are (1) inescapable and authoritative and objectively reason-providing or (2) they are inescapable and anti-authoritative and objectively reason-denying or (3) they are inescapable and a-authoritative and neither objectively reason-providing nor objectively reason-denying. My own position sides with (3), denying simultaneously that we have

reasons to be agents and that we have reasons not to be agents. The complex argument of the chapter does not support (3), of course, as it was never intended to do so. Its only purpose was to demonstrate the compatibility of the constitutive programme with radical meta-normative anti-realism, thereby showing its unavailability for being recruited to support let alone ground meta-normative authority. The alleged inescapability of agency-constitutive norms would have many implications, in particular with regard to motivation. In conjunction with normative premises it would even have normative implications, but then, that is true of everything. What is constitutive of agency and what is inescapable does not explain, ground, or account for what is authoritative and objectively reason-providing.

Chapter 5

Practical Deliberation Without Authority

5.1. Introduction

This chapter takes substantial steps towards developing a positive account of practical deliberation consistent with radical meta-normative skepticism, naming the resulting view *revisionary subjectivism*. In his treatment of the arch-subjectivist, Protagoras, together with his doctrine of *man being the measure of all things*, Richard Taylor (2000:56) provides an unflattering yet fairly accurate characterization of the view: “all moral distinctions are in the last analysis of human origin, such that nothing is in truth really right or wrong, just or unjust, but is only such within the framework of this or that artificial set of conventions.” Generalizing Taylor’s term *moral* to *normative* and this is roughly the position I defend. Notice revisionary subjectivism differs from traditional subjectivism, the approach of classical response-dependency, which holds, following

Spinoza, that “it is clear ... that we do not endeavor, will, seek after or desire because we judge a thing to be good. On the contrary, we judge a thing to be good because we endeavor, will, seek after and desire it.” (Spinoza, *Ethics* III.9 Scholium) Traditional subjectivism offers an affirmative view of what is good, right, or reason-providing, and why, intentionally placing the position in competition with alternative (objectivist) normative and meta-normative views, offering different answers to what fundamentally remain traditional normative questions. Revisionary subjectivism, in contrast, is a position of retreat, intentionally changing the subject matter, offering answers to rather different questions, and abandoning the normative project altogether.

Yet despite its retreat and according treatment of practical deliberation as a non-normative project, revisionary subjectivism aspires to save (enough of) the deliberative phenomena. The question is whether the skeptic qua subjectivist can pull this off. A major reservation about the global denial of all normative reasons stems from implications it is presumed to have for the possibility of individual and collective deliberation. The worry is that by abandoning the normative project the skeptic has also effectively abandoned the deliberative project, that, in the accusation of one author, the subjectivist, by necessity, must “keep[] very quiet about how little he is able to salvage from the wreck.” (Brandon 1980:98) Yet we are committed to the sensible nature of practical thought about how to live and act. Should the denial of normative reasons threaten that commitment, greater by comparison than our confidence in meta-normative skepticism, this result could handily be recruited to support meta-normative realism. With regard to ethical thought David Brink writes (1989:173): “If ... rejection of moral realism would undermine the nature of existing normative practices and beliefs, then the

metaphysical queerness of moral realism may seem a small price to pay to preserve these normative practices and beliefs.” The argument about reasons thus remains incomplete until we arrive at an assessment of the skeptic’s ability to accommodate enough of practical deliberation.

In response to this worry, hoping to save most of the phenomena and to salvage *a lot* from the wreck, the revisionary subjectivist – *subjectivist* henceforth – invents the internal stance, the idea of deliberation proceeding from *within*. Thomas Nagel captures the essentials of the idea nicely, together with the relevant dividing line between what the subjectivist desires to hold onto and what he is prepared to let go (1979:196-197):

There is a way of considering human pursuits from within life, which allows justification of some activities in terms of others, but does not permit us to question the significance of the whole thing, unless we are asking, from within life, whether the allocation of energy or attention to different segments of it makes sense in virtue of their relative importance. This view comes under challenge from a position that regards life in detachment from specific or general human purposes. People, and oneself in particular, are perceived as having no significance, and absurd because they seem to accord their lives great importance in action, even though they can also appreciate a broader point of view from which they have no importance.

Each of the two points of view claims priority. The internal view asks, what is the importance for individual life of insignificance from an external point of view? Life is lived from inside, and issues of significance are significant only if they can be raised from inside. It therefore does not matter that from a point of view outside my life, my life does not matter.

Revisionary subjectivism results from the simultaneous adoption of both viewpoints. Nagel's passage is characteristically rich, getting right to the crux of the matter, and I wish to restrict myself to two comments. First, while the last paragraph articulates very forcefully the general subjectivist spirit, the subjectivist should not kid himself, as do some non-revisionary subjectivists, to regard the indeed questionable internal significance of external insignificance (or external *significance* for that matter), the failure of recovering external significance in terms of internal significance, to carry much dialectical force against the objectivist. What the objectivist is concerned about is not internal significance, but something else, and *its* significance is henceforth not contingent upon a type of significance which is altogether different (indeed is different *by definition*). The objectivist ought to be entirely unmoved by all this. Neither side gives or gains any ground here.

Second, when Nagel charges life with absurdity "because [people] accord their lives great importance in action, even though they can also appreciate a broader point of view from which they have no importance," he introduces a problem his statement may not express optimally, potentially representing the problem as too escapable, apparently invoking an equivocation of two different senses of importance, a first sense which people do accord to their lives and which presumably is the internal sense, and a second sense which people apply from a broader point of view and which presumably is the external sense. One cannot keep both senses constant and yet generate the absurdity: If both senses are internal, then even from a broader view our lives have importance, hence no absurdity. If both senses are external, then our lives either have or have not importance, regardless of a boarder or narrower viewpoint. The non-revisionary

subjectivist believes they do not have that kind of importance but contests that people do accord their lives with it, while the objectivist affirms that people do, successfully, accord their lives with that kind of importance, and again whoever is right we miss the absurdity.

What is Nagel after, then? The answer must be this: both senses of importance have profound resonance for us; we are anything but indifferent towards external significance. A concern for external significance can be recovered from within, deeply entangled with our internal-significance based concerns. “This dominance [of the external viewpoint] is not imposed from outside, but derives from the intrinsic appeal of impersonality to individual reflection. Life seems absurd because it seems absurd to *oneself*, taking up a point of view that is both natural and appealing.” (1979:197) A sensitivity to the external viewpoint together with a recognition that subjectivism cannot accommodate it sets apart revisionary subjectivism from traditional subjectivism. Revisionary subjectivism is a form of discontented subjectivism, a position of graded pessimism. Nagel’s diagnosis carves out precisely the kind of position I am advocating in conjunction with the challenge it faces: affirming internal significance, denying external significance while being cognizant of its profound resonance, recognizing further that the second component threatens to destabilize the first component, in consequence defining the subjectivist’s task as seeking to stabilize internal significance in the absence of external significance.

The revisionary subjectivist should be the first to admit his account won’t deliver it all. Thus the question: how high should the subjectivist aim? His objective is to develop a success theory rather than an error or failure theory of deliberation, but success and

failure come in degrees. How much of the phenomena should he attempt to save? How much of deliberation should he hope his account will be capable of accommodating? What range of practical question should he hope will find deliberative answers, what range of common expectations and beliefs about deliberation should he hope will come out true, in line with his account? This is no obvious question. As a dialectical matter, his account will better be none of total deliberative failure. Deliberation happens, and as already mentioned we are quite confident it happens at least occasionally with success. Yet at the same time his account will better be none of total deliberative success either. If all we ever wanted from deliberation found its place in a comprehensively reason-denying framework, one would start to wonder what precisely it was the framework denied in the first place. *Something's gotta give* in a skeptical meta-normative account of practical deliberation. Generally speaking, the more the skeptic can accommodate, the less interesting becomes his meta-normative denial. The less he can accommodate, the less attractive becomes his overall package. Perhaps what he should shoot for is roughly half? That seems roughly right. The objective of subjectivist, then, is to present half-a-success, half-an-error theory of practical deliberation.

Can we be more precise? Perhaps the skeptic is entitled to an upper and a lower limit of how much we should expect him capable of accommodating? With regard to the lower limit, he might reason thus: if meta-normative skepticism is true, and deliberation at least occasionally happens with success, then the best subjectivist account must be capable of accommodating *at least that much*. This would protect subjectivism from falling into total abyss, from salvaging too little from the wreck. Unfortunately, the reasoning does not quite go through. It would be simply question-begging in this context

to assume skepticism is true. To the subjectivist's style of argument, broadly following a modus ponens patterns, the realist could retort another style of argument, broadly following a modus tollens pattern: deliberation at least occasionally happens with success, subjectivism cannot even accommodate that much, and hence subjectivism is false. Moreover, most believe skepticism to be false, and consequently even if skepticism was true, the success we find in deliberation might hinge on our illusionary belief. Since the subjectivist account of deliberation aspires to rely on no deceptions, it may still fail to accommodate as much deliberation as we currently observe works; if so, the subjectivist account would lose out against the fictionalist – the subject of the first section.

With regard to the upper limit, the skeptic might reason thus: deliberation quite frequently appears to fall short of what people expect it to achieve. Practical stalemates abound, myriad of practical questions remain unresolved, intractable ethical disputes stay intractable, etc. The subjectivist should certainly not hope to make deliberation appear more of a success story than it actually is; moreover, should subjectivism entail limitations of what deliberation can accomplish, limitations roughly matching the limitations we find in reality, would subjectivism not win some plausibility points? Unfortunately, the reasoning does not quite go through either. Deliberation is hard even if realism prevails. That many practical questions remain unsettled need not indicate systematic deliberative limitations, but may only indicate that we have not tried hard enough, that we are biased, that irrelevant considerations unduly influence the process, that we are in the grip of framing effects, nudging effects, swaying effects, or any other of the myriad well documented sorts of cognitive shortcomings. Not much greater precision can be provided regarding the subjectivist's objective. *Roughly half* thus stands.

I should also note that his endeavor to revise the deliberative story the subjectivist is not venturing out into entirely new territory. His position is rather similar if not identical in crucial aspects to various forms e.g. expressivism, quasi-realism, cognitive expressivism, and in large measures the subjectivist can adopt what their proponents have developed vis-à-vis deliberation (e.g. Gibbard 1990 & 2003, Blackburn 1998, Timmons 1999), despite all protests of expressivists against such association. For sure, as expressivists correctly reiterate, there is a semantic difference between an expression of an attitude and a reporting of an attitude, traditionally aligning the semantic agenda of expressivism with the former and that of subjectivism with the latter (an alignment I consider rather optional). Yet even if this is so, the question remains whether that distinction matters to the larger issues. For each position, all comes ultimately down to contingently held attitudes void of objective authority, and as Richard Joyce (2006:58) beautifully states, despite all technical differences, both analyses pretty much concur in what informational message is conveyed as well as what significance (or lack thereof) they seem to carry:

Expressing one's feelings and *reporting* one's feelings are different linguistic activities – on a roller coaster, screaming is not the same as stating calmly “I feel very excited now” – yet both provide one's audience with pretty much the same information about one's inner states. But simply to report the presence of a feeling is never, in itself, to provide one's audience with a practical consideration. Suppose that Roger is a vehement anti-hunting activist. As a group of fox hunters trot by, Roger asserts to them: “Your activities arouse a feeling of disapproval in me.” We could hardly fault the fox hunters if they responded with a perplexed “Yes, but so what?” But then much the same response would be

reasonable if Roger, instead of reporting the presence of disapproval, *expressed* his disapproval by yelling “Boo to fox hunting!”

Another companion in developing a subjectivist account of deliberation is Michael Bratman together with his extensive planning theory of intentions, which I would be happy to co-opt without drastic changes (this is not to suggest Bratman is a subjectivist; only that what he says can, in large measures, be incorporated by one). David Velleman’s intelligibility-based account of practical reasoning, subject of the previous chapter, provides another source of inspiration, yet the details of his account give me greater pause than Bratman’s, as we saw there. I mention this in order not to create the false impression that I am setting out to entirely reinvent the wheel, that in fact there is a larger research tradition the subjectivist’s revisionary account is part of. That said, the existence of a larger research program is far from a guarantee for its success, leaving many grave problems unresolved, occasionally even unaddressed and more often underestimated by a good number of fellow subjectivists.

The chapter is sprinkled with first-personal observations, presented in the hope of striking a cord with the reader, enticing him to come up with similar observations in himself. This approach is prone to cause some puzzlement, however, and hence a word on methodology seems in order. My objective is to show the meta-normative skeptic or subjectivist capable of accommodating a decent range of the deliberative phenomena. Yet what the skeptic or subjectivist is certainly not capable of is to prescribe what deliberation, fundamentally, ought to be like. What he is capable is to draw a picture of what successful deliberation could be like, pointing towards elements within that picture explaining what we seek to explain, in particular, pointing towards elements explaining

how certain deliberative steps can be said appropriate or inappropriate within certain deliberative context. What the subjectivist proposes is *a* case study – a study of *one* case – of successful deliberation consistent with meta-normative skepticism; to that end the subjectivist best volunteers himself as the case to be studied. Generally speaking, in light of the skeptic's affirmation of standards of correctness in conjunction with his denial of correct standards, description and prescription ultimately converge at some point in his overall approach. Normative inquiry culminates in self-understanding (in a form of psychoanalysis, a cynic might add).

Daniel Dennett formulates a similar idea, and borrowing a term from Wertheimer, calls the underlying methodological principle the *Factunorm Principle*. Here is what he says, with regard to norms of rationality (1987:98, emphasis added):

When one leans on our pre-theoretical concept of rationality, one relies on our shared intuitions – when they are shared, of course – about what makes sense. What else, in the end, could one rely on? *When considering what we ought to do, our reflections lead us eventually to a consideration of what we in fact do*; this is inescapable, for a catalogue of our considered intuitive judgments on what we ought to do is both a compendium of what we do think, and a shining example (by our lights – what else?) of how we ought to think.

Dennett then proceeds to quote Wertheimer (1974:110-11):

Thus, what and how we do think is evidence for the principles of rationality, what and how we ought to think. This itself is a methodological principle of rationality; call it the Factunorm Principle. We are (implicitly) accepting the Factunorm Principle whenever we try to determine what or how we ought to think. For we must, in that very

attempt, think. And unless we can think that what and how we do think there is correct – and thus is evidence for what and how we ought to think – we cannot determine what or how we ought to think

Let me provide a brief overview of the chapter. It starts, in section 1, with a discussion of the contest between fictionalism and revisionism, attempting to show revisionism to prevail. The argument mounts an internal critique of fictionalism, seeking to entice the fictionalist to cross over to the revisionist side. After this the chapter presents, in section 2, a succinct formulation of the challenge subjectivism faces and must overcome, namely the above mentioned worry that without a normative account there is no adequate deliberative account. This is followed, in section 3, by my proposed solution schema. Next, in section 4, is a discussion of the question *So it is all about motivation?* This sets the stage for introducing, in section 5, the true hero of this chapter, desire, or concern. For the subjectivist concerns are what get deliberation going. Attending to concerns more closely, in section 6, is going to reveal an astonishing degree of complexity and richness. Moreover, we are going to find, in section 7, that concerns contain an element of appraisal, which initially appears hard for the subjectivist to capture. In section 8 I detail the subjectivist's solution for how to capture appraisal. The final section 9 presents a case study of concern – the concern for recognition – designed to illustrate how the subjectivist can capture appraisal.

This discussion has largely crowded out an originally planned detailed discussion of norms and commitments, additional building blocks of the subjectivist's deliberative account; I do not regret the result. Practical deliberation is complex and untidy, a fact missed by many philosophical treatments by impatiently hurrying over its details. There

is a lot about concerns not fitting prevalent philosophical conceptions, quite often seemingly owed more to a desire for generating neat and simple theories than to a close observation of the phenomenon. I wholeheartedly share Velleman complaint that “[t]he agent portrayed in much philosophy of action is a square.” (2000:99) There obviously are many more sides to agency and deliberation than just these plain old four, and it would be quite some progress if my account of deliberation was able to represent agency at least as an octagon. Ambitions for simple and smooth theories have attempted to mold agency into a simple and smooth phenomenon, ignoring the numerous sharp edges, rifts, and ruptures; smoothing it out, at worst turning agency even into a circle, betraying a common yet astonishing unwillingness to come to terms with who we are.

5.2. Revisionism, Not Fictionalism

Fictionalism enjoys increasing attention as a practically beneficial solution offered after we have become convinced of the falsity of certain practical domains. Moral fictionalism is the practical case most thoroughly discussed in the literature and thus shall be the prime subject of this section, facilitating exposition without limiting the general applicability of my conclusion. What we learn about moral fictionalism can be extended to other forms of practical meta-normative fictionalism. Richard Joyce sets the scene for fictionalism’s appearance (2005:288):

Let us just say when morality is removed from the picture, what is practically called for is a matter of a cost-benefit analysis, where the costs and benefits can be understood liberally as preference satisfaction. By asking what *we* ought to do I am asking how a *group* of persons, who

share a variety of broad interests, projects, ends – and who have come to the realization that morality is a bankrupt theory – might best carry on.

The promise fictionalism offers is that we may retain the benefits of knowingly discredited practices by downgrading our epistemic attitudes towards them from *believing* to *make-believing*. The promise is pragmatic. Fictions cannot be epistemically justified, for they are plainly and literally false, and thus the only kind of justification left is pragmatic in nature. The central motivation for fictionalism is the claim that the fictionalist stance importantly contributes to our ability to retain the benefits of the discredited practice, where this is read as a comparative claim: We are better positioned to retain the benefits by make-believing rather than by not make-believing in the discredited practice. The relevant comparison-class for fictionalism includes revisionism and eliminativism. I am not particularly attracted to eliminativism, the abolishment of the discredited practice tout court, and thus I am happy to provisionally grant fictionalism to win that contest.

Revisionism fares considerably better, however. Should it be possible to revise the discredited practice, by removing what is erroneous while keeping what is beneficial, in such a way as to sustain what we want, the twisted allegiance to the original practice already recognized as flawed that characterizes fictionalism becomes entirely moot. Problems pertaining to the fictionalist attitude are myriad and well documented (Nolan 2005). I shall not reiterate them here. The comparatively weak assumption I am working with is simply that everything being equal, practices proceeding on the basis of truthful attitudes are preferable to ones proceeding on the basis of commonly known false attitudes. Here, then, is gist of my argument. It seems plausible to suppose that the

benefits the fictional attitude is capable of affording is owed to the retainable elements of the corresponding practice. If so, there always is a revisionary alternative available which “shrinks” the original practice to precisely those retainable elements and which has the advantage of clear-headedly dispensing with the barely-stable fictionalist attitude. If so, we better stop *make-believing* in what does not work and start *believing* in what does work.

Joyce’s treatment of fictionalism together with the pragmatic nature of the relevant test case is mostly admirably clear. It is very surprising to learn, then, that revisionism does not feature as a relevant competitor in his argument, where fictionalism supposedly faces only one single competitor, namely eliminativism. Joyce writes (2005:299): “Two options remain as contenders in the cost-benefit analysis: abolitionism (or we may call it ‘eliminativism’) and fictionalism. For fictionalism to be viable it must win this pragmatic comparison,” speaking of a “two horse race” on the same page, and again emphasizing later on that “the only comparison in which we are interested is between fictionalism and abolitionism.” (2005:302) As I already said, I believe fictionalism is well positioned to win that contest. The glaring omission of revisionism is rather unexpected, however, especially since it has always formed a prominent part of error-theoretic discussions. If morality is bankrupt, instead of electing to maintain morality as a fiction, we may alternatively elect to switch our allegiance from morality to shmorality, a system much like morality except for omitting its errors, as true believers of an adequately revised scheme rather than make-believers of some flawed scheme.

To focus the comparison between fictionalism and revisionism, we need to be clear on what kind of fictionalism is at issue. Three points bear attention. First, we are not talking about deception. All practitioners of the fiction are assumed privy to the character of their fictionalist stance. This is not an upper-house version where the folk are kept from the truth by some epistemically privileged elite. Joyce calls such a version “propagandism” where “*some* people may be ‘in the know’ about the moral error theory while, for the greater good, keeping it quiet and encouraging the *hoi polloi* to continue with their sincere (false) moral beliefs.” (2005:299) I share Joyce’ assessment that propagandism is generally a bad idea. The most obvious shortcoming of such a strategy stems from the destabilizing effects it is sure to have on public discourse and critical inquiry for truth. And neither are elitists ever to be trusted, as they are likely to substitute their own interests for the interests of those who they are supposed to look after. It therefore appears legitimate to place propagandism beyond contention and have it take no part in our present comparison. It is true but irrelevant, then, that non-transparent deceptions often are capable of conferring all sorts of practical effects, some beneficial some not. That there usually are revisionary alternatives to deceptive fictions with comparable practical effects is dubious and not what I am arguing for. What I contest applies only to perfectly transparent fictions.

Secondly, fictionalists are *convinced* that the corresponding practices are *flawed*. This reminds us of two sub-points. Number one, fictionalists are not agnostic harboring at most some mild doubts about the success of the corresponding practice. It appears reasonable to take a conservative attitude towards what works and has not been proven

deficient, even if the epistemic case in support is comparable to the case against. Fictionalists are not conservative agnostics in this regard, but ardent deniers.

Number two, the flawed practical domain is really considered flawed, as falling short in its original function or purpose. One is not a fictionalist when one believes the point of some practice consists in something other than truth. In fact, one cannot even be a fictionalist in such an instance. One cannot be a fictionalist about fictions, for instance, since their purpose is not truth, but entertainment or some such, raising some questions about the appropriateness of the term *fictionalism* itself. The pertinent attitude towards fictions is not make-believing or pretend-believing, but entertaining and imagining, cognitive attitudes that do not constitutively aim at what's true as we learned from the previous chapter, and consequently suffer no defect if shown to be literally false. Indeed, one might suspect that with an appropriate provision of cognitive attitudes other than belief – including the intriguing attitude of *Alief* as characterized by Tamar Gendler (2008) – the case for inevitably shifting to make-believing after believing has been taken off the table turns out somewhat muted. This is relevant especially in connection with Joyce's important insight that many of our moral attitudes come in the form of pre-reflective commitments, viscerally anchored resolutions to adopt a moral point of view, usually steadily habituated from early childhood on, and which are unlikely affected by error-theoretic meta-ethical considerations, since they do not incorporate the kind of abstract meta-ethical presuppositions that could be contradicted by the very same considerations. The issue of commitment shall occupy us more thoroughly in later sections; the suggestion now is that commitment and especially pre-commitment (Jon

Elster 1985) may involve cognitive attitudes other than belief, in which case no downgrading to make-belief is needed.

Thirdly, fictionalist must take a discriminatory attitude towards flawed practices, where they identify some but not all elements as defective. The defective elements must be suitably central so that they infect the practice as a whole. But at the same time, there must be elements of the practice that are non-defective as well. A practice completely and thoroughly bankrupt is an unlikely candidate to feature in any useful fiction. There must be many things that are true of the practice. To begin with, it needs to have the desired effects, precisely the kind of effects the fictionalist wishes to preserve through the fictionalist stance in the first place.

Turning finally to my general recipe for eliding fictionalism, I must start with a brief cautionary note. The contours of my revisionist strategy tailored towards rendering fictionalism superfluous are admittedly programmatic – the same is currently true of the fictionalist programme – and I must ask the reader for patience as additional details concerning the revisionary scheme I favor shall emerge as we proceed in the chapter. Moreover, the question of which competitor has the upper hand pragmatically speaking involves thorny empirical issues concerning the nature of human psychology and its interaction with make-belief for which we currently lack conclusive data; nor do we have cause to expect the kind of large scale social experiment required to garner them to materialize anytime soon. All we can do at this stage is to speculate about it, which of course is always fun to do.

To facilitate exposition, I propose we focus on a simple example, namely the moral status of pain. Let's first recap the general situation as both fictionalists and revisionists see it: We used to be caught up in the (deceptive) myth that pain is such as to render minimizing measures toward it (categorically) *appropriate* or *obligatory*, until we learned that this is not so, and that instead we only dislike it and seek to minimize it. In response the fictionalist proposes that we falsely *make-believe* that pain has that peculiar feature so as to better enable ourselves to minimize it. By doing so the fictionalists adopts the kind of discriminatory attitude toward the original practice mentioned above where some things are erroneous, motivating the fiction, whereas other things are not erroneous, motivating the assessment that the practice is worth preserving in the mode of a fiction conferring true benefits. What is erroneous and what needs fictionalist treatment is the supposition that pain has the moral status of rendering a minimizing response towards it (categorically) *appropriate* or *obligatory*. What is not erroneous, in contrast, is that we harbor a strongly disvaluing attitude towards pain, and that we are committed to fostering schemes that help minimizing it. If we cannot count on some reasonable consensus to minimize pain the fictionalist programme collapses before it has truly started. After all, to the question why he chooses to retain some flawed scheme as a fiction, the fictionalist points out that doing so gives us something we truly value. Now *that* we value what we value is no fiction, and neither is it a fiction that engaging in a certain practice helps realizing what we value. The value of the fictionalist stance entirely traces the value of what we care about.

It is precisely at this point that the revisionist enters the scene, asking the following question: Why not just understand what we are doing as *doing-that-which-*

helps-realizing-what-we-value? Why not make do with the clear-headed appreciation of what we are doing which, as we assume, we already muster, and stick with that? What the revisionist proposes in contrast is that we dispense with any make-belief altogether, and simply agree to minimize pain which we so strongly dislike. The claim I am going to support in the ensuing paragraphs is that the revisionist should be in no worse position to achieve this goal than the fictionalist his, once we appreciate the complexity of the material the fictionalist is already committed to for his fiction to work.

To better see why, we need to consider how highly disciplined and focused the fictionalist discourse must be in order to stand any chances of procuring real benefits. For starters, no good comes from make-believing that what matters are suffering, destruction, and despair. History contains plenty of moral ideals, previous drafts of the morality script as it were, that poorly comport with what we care about. The moral code for medieval monks is particularly unbecoming, prohibiting them, in recognizably moral fashion, from securing even the most minimally adequate provision of food, sleep, comfort, social exchange, means of expression and leisure, maximizing rather than minimizing pain. For fictions to be doing us any good, they must exhibit a close tie to what really matters to us. In light of this, the practitioners to the fiction are well advised to retain a clear grasp of what really matters to them, making sure they understand what is the tool (the fiction) and what is the original purpose (facilitating their concerns). It is well documented how collective practices, including fictions, are keen to develop a life of their own, possibly striving to emancipate themselves from their initial purpose, possibly straying away from their original motivation of helping us to get what we want. There always is plenty of internal and fiction-specific pressure in the direction of certain modifications and

narrative enhancements that produce better and more enthralling fictions – just recall the common tendency to embellish even the most mundane reports – yet which present a considerable risk of diminishing the fiction’s capacity to serve our ends. At precisely these moments we need to consult what it is that really matters to us, and what is only fictitious, and employ that grasp as a corrective device for readjusting the fiction. The fiction must always be kept on a short leash, never forgetting who is master and who servant.

Moreover, the fiction must be highly disciplined. There must be tight standards regulating what may and may not be asserted within the adopted fiction, which moves are and which are not admissible within the game of fiction in force. Though fictional, it cannot be that anything goes. It cannot be admissible, for instance, to assign significance to pain on Mondays but not on Tuesdays, or to pain suffered by men but not by women, or to pain caused by people as opposed to circumstances. There are in fact myriad alternative ways of treating the significance of pain – alternative fictional pain-status-functions as it were – concerning such contested matters as whether or not it is permissible for persons to privilege their own pain, to discount deserved pain, to balance pain against other concerns in certain ways, and so on. Some standards of consistency and coherence must be paramount, but standards of consistency and coherence alone don’t even begin to capture all of the regiment required to permit the fiction to function. In the moral domain, we would predict the regiment we ultimately settle upon to be as complex as the best moral theory has it. Of course, which theory is best is, according to the fictionalist, to be pragmatically determined. The question of which fictionalist regiment to implement cannot be settled by appealing to which regiment is the correct

one, since the guiding assumption is that nothing literally possesses the relevant status in the first place, but instead must be settled in terms of their capacity to help realizing what we care about.

Now suppose we can sum up all norms governing our fiction in *N*, and suppose we can evaluate *N* in terms of how well it serves our ends *E*. *N* might be formalized, in part or as a whole, in terms of various bi-conditionals, such as “in situation *S*, it is correct to assign to complex *C* the fictional status of rendering response *R* (categorically) appropriate or obligatory.” Within the fiction, it might be correct to assign this status to undeserved yet not deserved pain, for instance. And what standard of fictional-correctness to implement would be determined with regard to their capacity to promote *E*. This two-prong approach is familiar enough, as it is employed by some forms of indirect consequentialism, or practice-based approaches as articulated by the early Rawls (1955). The upshot for our pragmatic comparison between fictionalism and revisionism will be this. With *N* and *E* in place, the revisionist has almost all he needs to advance his alternative revised practice: as a practice in the service of *E* governed by *N*. This is not to deny, of course, that there will not be elements he needs to modify. In particular, the revisionist must propose a replacement status to be assigned instead of the fictionalist status. Speaking of pain, the erroneous, yet within-the-fiction tolerated, assumption that pain has the status of rendering minimizing measures appropriate must be characterized in terms of some different status, possibly one assigning pain the status-function of calling upon certain responses according to the norms governing the revised practice. The revisionist counterproposal would likewise proceed, in part or as a whole, in terms of various bi-conditionals: “in situation *S*, it is correct to assign to complex *C* the literal

status of rendering response R (categorically) appropriate or obligatory *according-to-our-chosen-practices-in-the-service-of-our-ends*.” The task of matching fictionalist bi-conditionals with revised literal bi-conditionals appears quite manageable, especially in light of the fact that nothing literally has the fictionalist status, but rather that it was entirely up to us to fictionally assign the relevant statuses in the first place, and that all this is common knowledge.

The smoothest way for the revisionist to proceed is probably by retaining the original normative vocabulary, only to supply a revisionary definition which preserves the literal truth of most of our original judgments. The original linguistic practice of calling pain *bad*, with the intended meaning of thereby assigning pain the status of rendering minimizing measures appropriate or obligatory, a falsehood the fictionalist proposes to retain in the modus of make-belief, is supplied a modified definition by the revisionist. What the revisionist might suggest, for instance, is that we redefine “bad” as denoting not what renders minimizing measures appropriate or obligatory, but as that which we agree to minimize, or as that which we are supposed to minimize *according-to-our-chosen-practices*. Notice how close the revisionist’s counterproposal is to what the fictionalist must already acknowledge. For the fictionalist, pain is assigned some fictional status according to the tight regiment governing the fiction in force. For the revisionist, pain is simply to be treated in certain ways according to the norms of the practices he recommends we adopt.

The present point, then, is this: once we appreciate how focused and disciplined the fiction must be in order to stand a chance of doing any good, we are already on our

way of granting much of what the revisionist needs, namely certain standards of assertability-conditions governing the revised practice which are not understood in terms of some direct correspondence to some normative reality. The fictionalist regiment, when suitably worked out, would offer great inspiration in this regard precisely because it is understood that the standards of fictional correctness are not determined by how well they track some normative reality. And the more we attend to the highly focused and disciplined regiment underwriting the fiction, the greater appears the contribution of it and the lesser the contribution of the fiction. The source of the benefits allegedly afforded by the fiction now appears to lie in the fact that we engage in a highly focused and disciplined discourse governed by some tight regiment. As the fictionalist contribution increasingly becomes marginalized, the revisionary alternative is to simply cut the fiction and to restrict morality to the regimented discourse in the service of human flourishing. What the revisionist proposes is that we focus our attention on what, according to the fictionalist, was really motivating the fiction in the first place, namely its capacity to serve our ends, and dispense with the fictionalist veneer that cannot fool anybody anyway.

Let me close this section with some further and quite intriguing speculations Joyce adduces in support of the prediction that fictions are capable of making a real psychological difference for non-believers-yet-make-believers. What distinctive psychological contribution does Joyce believe the fictionalist stance holds on offer? His answer has to do with temptation. To answer the question more thoroughly, Joyce proposes to first examine the value of moral belief, and in particular in distinction to the role of other factors. Joyce concedes, quite plausibly, that a good deal of our reciprocal,

cooperative and other-regarding attitudes do not depend on moral belief, and hence need little aid from fictionalism after the departure of morality. With some fusion between Hume and Hobbes, combining elements of enlightened self-interest, modest other-regarding sympathy and the enforcement of socially beneficial codes of conduct, one should probably be capable to cover a considerable range of the desired effects associated with morality. Yet however refined this post-moral and prudence-based social system may be like, it cannot deliver it all, since there is temptation. Joyce (2005:301) writes:

[An] important value of moral beliefs is that they function as ... an expedient: supplementing and reinforcing the outputs of prudential reasoning. When a person believes that the valued action is *morally* required – that it must be performed whether he likes it or not – then the possibilities of rationalization diminish. If a person believes the action to be required by an authority from which he cannot escape, if he imbues it with a ‘must-be-doneness’ ..., if he believes that in not performing he will not merely frustrate himself, but will become reprehensible and deserving of disapprobation – then he is more likely to perform the action.

Moral belief features as a kind of conversation-stopper, as an element that remove certain practical options from further consideration, an idea that has found traction with many writers, including Dennett. Joyce (2006:164-165) describes the idea of a conversation-stopper nicely as

[C]onsiderations that can appear in a personal or interpersonal decision process in order to bring the procedure to an effective terminus, preventing the mechanisms from spiraling endlessly in seeking further justifications, infinitely mulling over further considerations. For Dennett, moral values, principles, and imperatives function as conversation-stoppers par excellence. An interesting feature of the view is that in order for our moral

judgments to fulfill this role effectively we must not be thinking of them as conversation-stoppers at the time of deploying them; we must, rather, think of them as expressing genuine, no-questions-asked, end-of-story, moral considerations.

Thus we get an explanation of the practical import of moral beliefs; in particular, we get an explanation for why it is not implausible to suppose that moral beliefs qua conversation-stoppers have the potential of combating unbecoming temptations by taking off the table what would otherwise require considerable effort to resist. Yet what we don't yet get is an explanation of the practical import of transparently false make-belief, and in particular an explanation of why false make-belief should have practically superior effects than revised and modified belief. The big question, in Joyce's own words, is "how can a fiction have the kind of practical impact – moreover, the kind of practical *authority* – that confers on moral belief its instrumental value?" How is it possible that "a mere fiction could or should have such practical influence in important real-life decisions"? (2005:301) The suggestion is that moral make-belief inherits some of the practical gravity that usually is attached with moral belief. How could that work?

Joyce considers the example of daily workout (2005:303):

Suppose I am determined to exercise regularly, after a lifetime of lethargy, but find myself succumbing to temptation. An effective strategy will be for me to lay down a strong and authoritative rule: *I must do fifty sit-ups every day, no less*. I am attempting to form a habit, and habits are formed – and, for the doggedly weak of will, maintained – by strictness and over compensation. Perhaps in truth it doesn't much matter that I do fifty sit-ups every day, so long as I do more-or-less fifty on most days. But by allowing myself the occasional lapse, by giving myself permission

sometimes to stray from the routine, I pave the way for akratic sabotage of my calculations – I threaten even my doing more-or-less fifty sit-ups on most days. I do better if I encourage myself to think in terms of fifty daily sit-ups as a non-negotiable value, as something I must do if I am ever to get fit.

And that last piece, namely the proposition maintaining the absolute necessity of fifty sit-ups, is transparently false yet supposedly best affirmed in the modus of make-belief, according to Joyce. What he describes is a real psychological phenomenon most of us are struggling with. What I hesitate to grant, however, is that any transparently false make-belief is needed or comparatively better positioned than some clear-headed alternative belief to ameliorate the present practical quagmire. In fact, it seems to me that what Joyce describes is precisely what we should believe, and all that we should believe (or make-believe), since it seems to deliver the same practical benefits. It is true that Joyce, determined to exercise, needs to lay down a strict law permitting no exceptions. But the attitude he has no need for is that of falsely make-believing that only fifty-sit-ups could serve his athletic purposes. Instead, when facing temptation, what he needs to remind himself of is that breaking the law considerably weakens his habits, compromising his future resolve to do any exercise, and that as a psychological matter, he must fully abide by his strict workout plan if he is to stand any chances of sticking to it. Nothing in that realization is mistaken, supposedly requiring a fiction to maintain it, but rather represents a clear-headed awareness of the psychological implications of what we do now for what we may do later as well as our general dispositions and habits. Placing our current choices in a larger psychological context, appreciating their long-term effects among other things, endows them with considerable gravitas, quite possibly outweighing

the kind of hallucinatory gravitas knowingly absent in reality and present only in our fictional fantasies.

5.3. The Challenge from Deliberative Indispensability

Return, then to the challenge that the abandonment of the normative project entails the abandonment of the deliberative project, the challenge that no account which denies normative reasons is capable of representing practical deliberation as anything other than deeply flawed, forfeiting the capacity of making sense of practical deliberation at the moment of denouncing normative reasons. The challenge, to put it in yet one more way, that the success of the deliberative project entails the existence of normative reasons and hence the falsity of meta-normative skepticism.

David Enoch (2007) has recently developed a powerful version of the challenge against meta-normative skepticism and in support of what he calls *robust normative realism*, which he defines as the view that *there are response-independent, non-natural, irreducibly normative truths, perfectly objective and universal ones, that when successful in our normative inquiries we discover rather than create or construct*. (2007:21) The core concept in Enoch's argument is *deliberative indispensability*, which he juxtaposes to explanatory indispensability. Consider the latter first. Why believe in electrons? The answer is that electrons feature in our best (and decent) scientific theories. Electrons seem explanatorily indispensable to science, and with abduction we are entitled to believe that there actually are electrons. Moreover, as Enoch observes, science is not an enterprise we could easily dispense with. Clearly not all enterprises justify ontological commitments in

those features that cannot be eliminated from them without undermining their very point. Indispensability in science carries ontological weight because it is part of an *explanatory project* that is itself “indispensable because it is one we cannot – and certainly ought not – fail to engage in, it is unavoidable for us; we are essentially explanatory creatures” Enoch concludes “With non-optional projects like the explanatory one, there is no real option of abandoning them. If something is indispensable for such a project, it seems belief is the only rational way to go.” (2007:33-34)

Turn then to deliberative indispensability. Enoch argues that (robustly understood) normative truths are indispensable for deliberation, another enterprise we have no choice but to take seriously, and that this grounds belief in such truths. He rejects the idea that normative truths are *explanatorily* indispensable for deliberation. The progression of thoughts and intentions that constitutes deliberation is part of the natural order, and for all we know robustly meta-normative truths are not required for its explanation. Yet Enoch believes explanatory indispensability is not the only basis on which we incur justified ontological commitments. The indispensability of features that provide point and purpose to deliberation can justify belief in them as well. After all, we are no less *essentially deliberative creatures* than we are essentially explanatory creatures. “We cannot and should not avoid asking ourselves what to do, what to believe, how to reason, what to care about. ... The deliberative project is not one we can opt out of, it is not optional for us.” (2007:34) Consequently, if deliberation makes sense only if there are reasons, and we must deliberate, it seems we are committed to reasons: “by deliberating you commit yourself to there being relevant reasons; if you also believe there aren’t any, you are being inconsistent ... and irrational.” (2007:38)

Enoch's emphasis on our own deliberative perspective is well placed. How could we altogether fail to take seriously our own deliberative engagement together with what it entails? We possibly could adopt an uncompromisingly cynical attitude towards the deliberative attempts of other people who we could deplore as victims of thoroughgoing deception and illusion. But it seems hard to imagine how we could adopt a similar attitude towards our own deliberation. Even when we entertain the terrifying possibility of being subjected to massive error, we must at least have some confidence in our ability to reason towards a sensible response to that possibility: whether to take it seriously, and in case we do, what to think and do then. In general, even when we contemplate radical challenges to the sensibility of our own deliberation, we must do so qua deliberators who cannot but take seriously at least some of their own deliberative inclinations. If Enoch could demonstrate that we would be incapable of recovering point and purpose in our own deliberative engagement without believing in (robustly) normative truths, we would be under considerable pressure to form that belief. Indeed, given the essentially deliberative creatures we are what more convincing case can *we* imagine for such truths?

Enoch imagines a student contemplating whether to stay in law school or whether instead to switch into philosophy. The decision is of some consequence and treated accordingly. Questions are coming up concerning his happiness in law versus philosophy, the prospects of succeeding as a lawyer versus as a philosopher, as well as the two careers' potential impact on political issues, and so on. And then "There remains the ultimate question. 'All things considered', you ask yourself, 'what makes best sense for me to do? When all is said and done, what should I do? What *shall* I do?' (2007:36). These are fair observations, and here is how Enoch (2007:36) thinks they support realism:

When engaging in this deliberation, when asking yourself these questions, you assume, so it seems to me, that they have answers. These answers may be very vague, allow for indeterminacy, and so on. But at the very least you assume that some possible answers to these questions are better than others. You try to find out what the (better) answers to these questions are, and how they interact so as to answer the arch-question, the one about what makes most sense for you to do. You are not trying to create these answers. ... When trying to make up your mind, it doesn't feel just like trying to make an arbitrary choice. ... Rather, it feels like trying to make the right choice. ... Making the decision is up to you. But which decision is the one it makes most sense for you to make is not. This is something you are trying to discover, not create.

Enoch contrasts deliberation with mere picking. Consider the situation where we have to choose between two equally decent brands of cereal. Enoch continues (2007:37):

We can just pick in the face of a known (or believed) absence of reasons. But we cannot, it seems, *deliberate* in the face of a believed absence of reasons. Knowing that there is no decision such that it makes most sense for us to make it, we cannot – not consistently... – try to make the decision it makes most sense for us to make. Deliberation – unlike mere picking – is an attempt to eliminate arbitrariness by discovering (normative) reasons, and it is impossible in a believed absence of such reasons to be discovered.

Various authors have articulated arguments quite similar to Enoch's. Bennett Helm, for instance, writes, putting as much emphasis on *correctness*, *reasons*, *discovery*, and *non-arbitrariness* as Enoch does (2001:13):

[T]here seems to be an element of objectivity in what values a person holds in that she can deliberate about them correctly or incorrectly. Deliberation is a matter of choosing for reasons, thereby making possible

the articulation of why one course of life is better than another, so that it is not intellectually arbitrary which values we choose. Hence through deliberation we can *discover* the values things really have and so the kind of person it is worth our being, potentially overcoming delusions or misunderstandings about ourselves. ... [O]ur choice of personal values and so our understanding of the kind of person it is worth each of us being is not arbitrary ... For we think it is possible to have better or worse reasons for such choices ... The demand for discovery and objectivity is a demand to rule out this kind of arbitrariness.

A good dozen lines later Helm continues (2001:14, emphasis added):

[A]n arbitrarily chosen life lacks the kind of grounding that *is precisely the point of our deliberation* about value.

Call this the argument from *correct-answers*. It asserts that when we deliberate we incur a commitment to normative reasons that are relevant to our deliberation. (Read the assertion with all relevant qualifications in place: that not every deliberative occasion entails a corresponding commitment, that we need not be consciously aware of those commitments, etc.) The argument contains two steps: deliberation requires that there are correct and incorrect answers, and correct and incorrect answers require normative reasons. It begins by capturing something important about the phenomenology of deliberation. Figuring out whether to eat at home or to go out is not entirely unlike figuring out whether the restaurant is still open. There must be something to deliberative answers that is not exhausted by the fact that they just happen to be the answers we choose. Otherwise deliberation indeed collapses into mere picking where no place is left for reasoned decision-making. Step one seems fine. What the subjectivist contests is step two. Correctness in deliberative answers does not require (robust) normative reasons and

in fact represents a phenomenon the subjectivist can accommodate, although he must construe *correctness* as less involving than Enoch makes it out to be. The question is whether he can offer an alternative account that still permits point and purpose to what we are doing when we are thinking about what to do.

5.4. The Outline of a Solution

The contours of the subjectivist's counterproposal are predictable, and in fact were already anticipated in the opening sections. Deliberation receives correct answers from *within* the web of our concerns and antecedent norm-commitments we bring to the deliberative task, the subjectivist maintains, accounting for deliberative correctness conditionally, relative to the norms we employ, while his opponent insists on an unconditional reading. We deliberate when, driven by our concerns and norm-commitments, we actively participate in the resolution of practical problems, including the selection and development of which norms to live by, searching for common ground for how to coordinate our individual and joint endeavors. The deliberative enterprise is not unconstrained and is answerable to standards of correctness as it is carried out within a tight web of norms that we do already accept, a web we continuously spin and expand. The subjectivist does not have the resources to accommodate deliberative correctness outside this web of norms, correctness simpliciter as it were, and thus what Charles Taylor (1977) called radical or strong evaluation, an assessment of our most fundamental commitments on the basis of some view from nowhere, is not an option the subjectivist can countenance; he only has the resources to accommodate deliberative correctness from

within, his only card to play, which he needs to play well, rendering plausible that what he can offer is indeed enough of what we expect from deliberation.

Consider the deliberative task of figuring out how to organize the order in which students present in class. Suppose we conclude that it is best done by lottery. How did we arrive at that deliberative solution? We found it to be comparatively fair, since it does not flow from the mere swiftness and vigor with which students voice their preference, and we also prize the mutual harmony this decision procedure engenders. We know students will be more comfortable with their slot if they can rest assured that no one was unduly favored. We have, in other words, worked out a solution to a practical problem against a backdrop of shared standards. What precise content the standards have is hard to pin down. It probably includes a commitment that everyone has an equal shot in choosing his favorite slot. The point is that in light of those standards, whatever their content, the decision we made is recognizably better than many alternatives, and thus can be considered correct in a more profound sense than merely being the lucky one we chose.

The subjectivist applies the same account to deliberation about norms and principles itself. Norms are addressed to us and we must determine whether to take seriously what is so addressed, and thus we, the unique consumers of norms, inevitably face the task of deciding which norms to empower with action-guiding potential and which not, resulting in a division structurally resembling the realists' division between norms with and without authority. There are no objectively correct norms and principles that hold unconditionally, the subjectivist maintains, but standards of correctness are built right into norms, and thus granting him some he may commence deliberation about

others. In the context of deliberation about principles we shall no less rely on a horizon of norm-commitments within which to navigate further norm-selection. Standing upon more fundamental norms we may judge less fundamental norms, and judge in a manner that permits for correct results – intending the fundamental/non-fundamental distinction to be drawn contextually, vis-à-vis concrete norm-choice situations, with no norm principally enjoying a more fundamental status than others. The norms employed to judge others are taken for granted in the same deliberative context, momentarily suspending the request for justification in terms of yet further norms, with the possibility of reopening the case at any time.

Should we, in the progression to further and further norms, reach rock bottom at some point, then we would also have reached the end of our deliberative story, where little is left than to mention that this is simply how we are, that this is our deliberative way of life. Norms underwriting our planning capacities may be such an instance, in their pressure towards means-end coordination, meshing of sub-plans, diachronic stability, etc. (Bratman 1999) It seems irrational to refuse to take the obvious means to an end of major importance (assuming further conditions are met), and the norm driving that judgment may simply be built right into our cognitive design, with no further deliberative *because* involved, characterizing (most of) us as (mostly) instrumental reasoners, and that's that. Such ultimate norms, should they exist, would probably bear out themselves, unless they are extremely poorly designed, representing a case of some rather trivial self-certification, for sure, but not a case of bootstrapping to authority, since authority is no posit of the subjectivist's approach. It is true, then, that the skeptical account entails some limitations concerning the scope of non-trivial deliberation, but this should be fine so

long as enough space is left for the familiar kinds of non-trivial deliberation. Enoch's challenge is to explain how there can be deliberation *at all* if the skeptic is right, and that we can explain, namely how there could be point and purpose to deliberation about where to go given where we are. The skeptic and the realist may disagree about the scope of non-trivial deliberation, but that's an issue for another occasion.

Commitment looms large in this deliberative approach, and I need to say more about it, especially how it interacts with norms. Start with a challenge the skeptic faces. The skeptic qua subjectivist intends to stay committed to all sorts of norms, yet realists may wonder whether the skeptic is not forced to mischaracterize the nature of commitment. Without any confidence in the authority of the relevant norms, realists may complain, the skeptic's intention to say committed to them anyway must emerge as an unsustainable fetish, a sort of rule-worship in its worst form. The skeptic's response to the challenge depends crucially upon emphasizing a gap between the purportedly practical question of what norms to commit to and the purportedly theoretical question of what properties norms have, yet what is commitment if not acceptance in virtue of recognized authority, the realist asks. To this the skeptic responds by contesting the underlying assumption wedding commitment and authority by suggesting an alternative interpretation where commitments are divorced from judgments on normative authority. Skeptics say, rather, that commitments are stable psychological dispositions endorsed by reflection, to follow some norm's dictates, incorporating a readiness and willingness to be guided by it without regret; perhaps also accompanied by a further disposition to find certain features salient, or to exclude certain options from consideration. Commitments

further involve some *preparedness to see to it that one's intention to engage* [in the relevant norm-practices] *persists*. (Calhoun 2009:615-622, Frankfurt 1988)

Turn, then, to what commitment adds to norms. When we commit to a norm, we endow it with a special kind of motivational force, the only force left after normative force has been removed from the picture, with the consequence that we empower the norm to assume an action-guiding capacity, as a map we steer by. Vis-à-vis norms met by no commitment we can still work out what they entail is to be done, as a purely intellectual exercise, as when we say: “Yes, this is what I am supposed to do according to etiquette,” but hasten to add: “but I couldn’t care less about etiquette.” What we then experience quite immediately is the relativity of the relevant *supposed-to-do* claim to a standard wanting of all motivational impetus, failing to resonate with us entirely. In committing to a norm, in contrast, we allow it to take a more active role in our deliberative economy. The relativity of the demands it issues is still there, of course, but becomes seemingly motivationally discharged. In conjunction to our cognitive capacity enabling us to represent norms, supplying conditions of correctness to deliberation, we also possess the motivational capacity enabling us to commit, supplying practical force to deliberation. Neither capacity can help perform the job of the other, and commitment, in particular, cannot add further correctness-conditions, given its purely motivational/dispositional character, and thus whatever correctness-conditions are operative in some deliberative context must be fully accounted for in terms of the norms employed.

My approach echoes what Mark Timmons (1999) has developed with regards to moral knowledge: “Things get interesting when we judge from within a moral outlook, a position from which genuine moral assertions and truth ascriptions to moral statements have a home, as it were. And it is from this morally engaged perspective that talk of moral knowledge becomes interesting and useful.” Timmons contrasts this engaged perspective with a detached perspective where “the proper thing to say about moral truth is that there is none. That is, from a strictly metaphysical perspective from which we are asking about what there really is, since there are no moral facts or properties, there is no moral truth.” (1999:244; cf. also Wright 1992:200-1, Blackburn 1998) This nicely sums it up. I too wish to maintain that from without our commitments – the metaphorical view from nowhere (Nagel 1986) – deliberation cannot get off the ground. If we ever could adopt such a non-perspective perspective, we would have no choice but to notice the thoroughgoing nihilism it engenders. Fortunately, there is no deliberative view from nowhere, and even if, per impossibile, there were such a view, there would be no reasons to privilege its results rather than those generated by our own viewpoint.

Return, finally, to Enoch’s student. He cares about happiness, success, and the prospects of making a difference in the political world, among other things. Should he stay in law or switch into philosophy? His deliberation will be anchored in his norm-commitments, concerns, and his factual estimates: what promises do the two careers hold for his life, and which of the two so projected life-paths should he choose? Perhaps success leans towards law, happiness towards philosophy? The question then is what is more important to him. Initially this may be hard to tell. But after some thought he might reason that without happiness success is unlikely to arise; and that success in law appears

attractive only due to features that are incompatible with his commitment to political equality: dominance and showing off riches. Moreover, he may fear the many compromises of an increasingly troubling nature that success in law would require. Would he not in the end find himself defending big oil and vindictive divorcees? In contrast, the hardships of finding a job in philosophy when he is young are easier to bear than the frustrations as a lawyer when he is old. All in all, he may conclude success loses and happiness and philosophy wins. A variety of practical norms will condition the deliberative process and explain why the deciding factors support his ultimate decision. Such as the principle that if, everything being equal, out of two comparable goals only one comports with a significant third, one should choose that one.

5.5. So It Is All About Motivation?

This question, presumably asked in a critical voice, raises a host of issues it is crucial for the subjectivist to get clear on. Coming to a full appreciation of the distinction between the normative and the motivational is perhaps the greatest achievement in late meta-ethics. Major parts of the literature dedicated to illuminating the normative have emerged, upon close examination, to be treatments of another subject altogether, the motivational, calling into doubt their capacity to contribute to their intended subject-matter. The distinction was essential to my skeptical and negative case before, and will be essential for my subjectivist and positive case that follows. It focused my critical arguments. When I denied normative reasons, I did not deny motives, or motivational reasons, considerations that prompt people to act as they do, considerations that do

deliberatively guide people in their choices and which, in retrospect, explain and render intelligible what they did. The assassin in the service of the mob has an entirely intelligible motive for his kill, namely to collect his recompense, in light of which we understand what we did; given his motives, targeting at the overzealous investigator made sense whereas targeting at the random bystander would not have. *Making sense* and *intelligibility* are here understood as notions referencing familiar and expected motivational patterns, thereby providing answers to various why-questions, such as why the assassin intentionally killed the investigator rather than the bystander. Motives, and purely motivational reasons, are entirely kosher in my view.

Yet motives aren't reasons. It is a different matter altogether whether the assassin *should* have formed and acted upon the motive, whether he had any *reasons* of doing so, reading *should* and *reasons* in the characteristically normative way, as referencing not just some contingent standard he or we collectively happen to bring to the situation – some mobster-etiquette, say, according to which he was required to murder as promised – but as describing his normative situation as it obtains independently of any such standards except standards that are truly authoritative. The difference is clear. That the consideration to collect recompense by carrying out the assassination prompts, in accordance with familiar motivational patterns, the assassin to kill is what renders it a motive. That the consideration makes it appropriate for him to kill, not merely *according-to-some-non-authoritative-mobster-standard*, is what would render it a reason. That there are reasons and normative situation thus characterized is precisely what I deny, hastening to reaffirm my readiness, of course, to join and support the call for the implementation of social standards prohibiting and disincentivizing assassination.

The normative/motivational distinction should also give us pause of speaking too confidently of *The Normative Question*, as if there was only one single and well defined question going under that name, one specific kind of question we raise when we ask “why be moral?” or “why be rational?” or “why be an agent?” or “why do what is smart and wise and conducive to happiness?” It is unclear whether there is one single subject matter all those instances of the question ask about. The question “Why do or be such-and-such” is in fact ambiguous, allowing for a normative as well as for a motivational reading. The question “why be moral,” for instance, is normative when it asks about reasons to be moral; it is motivational when raised so as to prompt a decision about whether to be moral. When raised in the normative mode, the question queries about what normative status doing the morally right thing enjoys. When raised in the motivational mode, it presents its subject with a decision-problem, where the subject pondering the question is plausibly torn, having not yet made up his mind as to whether to be moral; a decision to form a motivationally efficacious state or intention is pending yet not concluded.

The two questions possibly bear all sorts of interesting relations to each other. According to one common picture, the motivational question is answered by answering the normative question. The motivational question asks for a decision, and a decision is selected on the basis of its merit. There are various complications about this neat picture. The present point is that the two questions are logically distinct, even if variously related, that they ask not about the same thing. Many authors ought to be more careful not to run them together, as for instance Korsgaard has made a habit of, so often speaking of the normative question when she plainly asks the motivational question. She writes, for

instance: "...if these arguments cannot *motivate* the reader to become a utilitarian then how can it show that utilitarianism is *normative*?" (1996:85, emphasis added)

The motivational/normative divide renders ambiguous most of our practical vocabulary. One particularly interesting instance concerns the notion of being *action-guiding*. Something may be said to be action-guiding in a normative sense in virtue of providing reasons for how one *should* act. Alternatively, something may be said to be action-guiding in a motivational sense in virtue of actually guiding how one *does* act. Most importantly, something may be said to be action-guiding in one sense without also being action-guiding in the other sense. There's also a third sense I should mention only to set it aside. Norms and practical standards can be said to be action guiding in the purely formal sense of having the right sort of content fit for speaking to practical questions of what to do. The standard *one is never permitted to tell a lie* is action-guiding in this formal sense whereas the proposition *people often do tell a lie* is not. Yet we know standards cannot acquire the kind of normative distinction associated with reasons-providingness solely in virtue of their form; they are action-guiding normatively speaking only to the extent to which what they specify that should be done must be done according to authoritative standards, not just according to some standard or other. Furthermore, standards need not have any motivational force in virtue of their form. Disregarding the formal mode of action-guidingness, we can thus distinguish between a normative mode where a standard is action-guiding just in case it provides reasons for what to do, and a motivational mode where a standard is action-guiding just in case is motivationally involved in how people actually decide to act.

To make the difference graphic, consider two maps of different sorts. Suppose a group of people is trying to get from point A to point B. There are two maps, and two routes. One map specifies the perfect route, the route they should take, whereas the other specifies the imperfect route, the route they should not take. Unfortunately, the first route is invisible, rendering it unusable by our traveling party. What they use instead is the second map. Now both maps clearly are action-guiding, but in a rather different way. The first guides by telling where they should go, albeit to no avail, the second guides by telling where they do go, albeit to no good. The perfect map, of course, would incorporate both aspects, telling where they should go and then also making them go there.

The example was intended to make the contrast between normative and motivational guides for action as stark as possible, speaking of an invisible map as opposed to a map merely currently unavailable, thereby undermining the very possibility that the invisible map, being action-guiding in the normative sense, could even be action-guiding in the motivational sense. This supposition is plainly incompatible with a prominent principle, namely that something can be normatively action-guiding only if it also retains the potential of being motivationally action-guiding. Expressed in terms of reasons, the principle holds that normative reasons must be considerations of the kind that people could possibly grasp and act upon, considerations, in other words, with motivational potential. The principle is false, however, and my suspicion is that only when reasons and motives are conflated with each other does the principle seem more plausible than it actually is. Whether certain considerations have the potential of motivating creatures of a certain kind depends first and foremost on the psychological

makeup of these creatures. Whether certain considerations favor certain responses presents another matter altogether. Once reasons and motives are set apart, it becomes hard to see why the kind of reasons that may apply to some creatures must yield to what the sorts of contingent psychological capacities they display permit them to grasp and act upon.

Notice the principle under consideration is distinct from another issue, namely whether creatures who grasp what their reasons are must, on pain of irrationality, be moved by them, representing a matter either trivial or false, depending on how the underlying notion of rationality is construed. Rather, what I contest is that reasons must be such that they can be grasped by those to whom they apply, and the force of my discontentment can be demonstrated quickly and with the help of a simple example. It is not implausible to suppose that, should there be reasons, dogs have reasons to ingest a sufficient supply of vitamins. There is something to be said in favor of that, namely that by ingesting a sufficient supply of vitamins dogs tend to have healthier and longer lives. Yet dogs cannot process thoughts about vitamins, due to their cognitive incapability to grasp and process the concept of vitamins, and consequently cannot act on these reasons – without the astonishing effect of thereby making them go away. The point generalizes. If there are reasons, it is not incoherent to suppose that some of them may be opaque to those to whom they apply due to cognitive limitations, specifying what they should do even though they cannot act on them. Applied to us, the way to describe the situation would be to say that there are some reasons for what we should do which we can understand and act upon and some which we cannot understand and hence cannot act upon. This is not entirely different from saying that there might be mathematical truths

we cannot understand, and the principle placing restriction of understanding and motivation on what reasons (ontologically) there might be appears just as absurd as the corresponding principle placing restrictions of understanding on what mathematical truths (ontologically) there might be.

To build on the example of invisible maps one last time, and expanding it by speaking of maps concerning how to act more broadly, suppose God laid out for us how to lead the perfect life and only forgot to use the right sort of print, or suppose alternatively he used a language we cannot decipher. His map is no less normatively action-guiding for that; unfortunately we simply are limited in our cognitive capacities to read it and consequently to be actually guided by it. Satan, in contrast, has perfectly mastered to decipher God's maps, with the inevitable effect of doing the exact opposite. His motivational guide for how to act perfectly matches his normative guide for how to act, only in reverse. It is essential for Satan to strictly keep normative and motivational guides for action apart in order to carry out his twisted and evil scheme. This completes my emphasis of the motivation/normative divide, and now I'd like to turn to what it means for my project as it is going to unfold in this chapter.

Returning to my own dialectical situation, I happily acknowledge that it was only due my insistence on the distinction between the normative and the motivational that I was able to permit my meta-normative denial focus and point, saving it from denying the obvious, namely that there are motives. Now I am stuck with it, including the implications this has for what I may say henceforth. Consider the following argument:

(1) The practical domain partitions, without remainder, into the normative and the motivational.

(2) The normative does not, and cannot, exist.

(3) One cannot truthfully talk about what does not exist.

(4) The motivational is all that remains to truthfully talk about in matters practical.

I accept the argument, as indeed I have but little choice. Having set the task of speaking of practical deliberation, aspiring to say many true things of it, all I can henceforth speak of is motivational in nature. If my meta-normative skepticism is correct, there is simply nothing else to talk about – nothing distinctively practical, anyway, for we shall certainly remain capable of truthfully talking about our world even after reasons have been removed from the picture. One thing in particular I wish to talk about are norms, what they demand, what they imply, and so on. The subject matter for the skeptic narrows to motivational matters as they interact with the world and norms. The only force I may countenance norms to have, besides the trivial formal force, is motivational force, after normative force, authority, has been dropped from the picture. The complaint all this must immediately arouse is this: aren't you thus changing the subject? Yes, I confess, guilty as charged, and in response plead that my hands are tied. Changing the subject from the normative to the motivational when setting out to developing a positive theory is inevitable for a meta-normative skeptic, the only way of avoiding massive falsehood given his assumptions. I fully understand the complaint of changing the subject, and I appreciate the force it retains for those who believe in the normative, and especially understand how it might serve as a remainder for those who, aspiring to illuminate the normative, have steered off target, as a call to refocus. Against

the meta-normative skeptic qua subjectivist the complaint is dialectically toothless, as pointless as criticizing the atheist for disengaging from theology.

I also grasp the low esteem of the motivational usually embodied in normative treatises. If one believes to have a choice of talking about the normative versus the motivational, the normative appears more exciting. And yet, the motivational is more exciting than often credited, especially the enormously sophisticated subspecies which we may call *rational motivation*, which we employ when we engage in complex long-term deliberation. The common depreciation of the motivational I seek to counter with an appreciation of the motivational, presenting it as a powerful capacity, particularly when employed for complex long-term planning purposes by ourselves. Once we acquire a better sense of what rational motivation involves, especially in its active mode, we shall become more sanguine about its prospects of accommodating most of what we take ourselves doing when deliberating. The force of this promise has to be assessed vis-à-vis the details of the ensuing account. There are various general obstacles that can and must be removed before we turn to that, though.

A first source of depreciating the motivational lies in the supposition that motivation is third-personal and concerns matters of explanation only whereas normative reason is first-personal and concerns matters of deliberation. This represents such an egregious misconception which must be addressed even before my motivation-based deliberative account is introduced. One author leading the charge is Jay Wallace who writes (2006: 66)

When we address the question of motivation, we typically focus on an action that has already been performed, and we think about the action from a distinctively third-person perspective. The aim is to arrive at a certain kind of explanation, taking the explanandum as a given. ... [I]n posing this question I am viewing my action from a point of view that is disconnected from my present capacities for agency... By contrast, the perspective within which normative reasons have their place is characteristically prospective, first-personal, and deliberative. It is the point of view from which I weigh the considerations for and against the various courses of action that are open to me, with the aim not of explaining something that has already occurred, but of resolving for myself the question of what is to be done and acting accordingly.

As a statement about reasons and motives most of what Wallace says in this passage is mistaken – a verdict that would be different had he attempted to describe the deliberative perspective more broadly. For instance, his tripartite claim that “the perspective within which normative reasons have their place is characteristically prospective, first-personal, and deliberative” is false on each count. We can query about normative situations of the past, concerning other people, and in a purely evaluative as opposed to deliberative context. The assessment that Caesar had plenty of reasons to secure better protection on his way to the Senate at the Ides of March plainly involves a perspective in which normative reasons have their place, yet is neither prospective, nor first-personal, nor deliberative.

My real source of disagreement, however, concerns Wallace contentions about motivation being third-personal and characteristically retrospective. Sometimes it is, sometimes it is not – the same is true of the normative. To see why Wallace is mistaken, I

suggest we momentarily redraw the normative/motivational distinction in terms of attraction, namely in terms of the contrast between what one *should* be attracted by (the normative) as opposed to what one *happens to be* attracted by (the motivational). First-personal and forward-looking deliberation can proceed in either terms. When deliberation is carried out in terms of normative reasons, we deliberate in terms of what we *should* be attracted by; when forward-looking deliberation is carried out in terms of motivation, we deliberate in terms of what we *are* attracted by. Think of a simple deliberative situation, the forward-looking question of where to have dinner tonight. The question – *Indian or Chinese tonight?* – prompts a decision-process which might very well unfold without making any reference to normative reasons, or to what one *should* be attracted by, but unfolds rather by considering merely what one *is* attracted by. Various attractions are brought to one's mind – initiating a mental search for suggestive culinary imagery, eventually conjuring up a savory thought of chicken-tikka-marsalla – which triggers ones little motivational whistles and wheels, ultimately culminating in the placement of the order at one local Indian restaurant.

A second source of deprecation of motivation stems from construing it too narrowly. We need to appreciate its richness and complexity. This plea for inclusiveness on my part is admittedly strategically motivated, since accepting limitations on the motivational will immediately limit its capacity to explain the rich range of deliberative phenomena I am setting out to explain. As a theoretic developer I am well advised not to restrict the territory to operate upon, except in so far as it concerns intrusions into normative lands. Still, the plea for inclusiveness is not merely a strategic matter but seems to be called for in its own right. Williams was keenly aware of the dangers of an unduly

narrow construal of motivation when, speaking of his famous motivational set S, he urged us (1981:105)

... to be more liberal than some theorists have been about the possible elements in S. I have discussed S primarily in terms of desires, and this term can be used, formally, for all elements in S. But this terminology may make one forget that S can contain such things as dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may abstractly be called, embodying commitments of the agent.

I believe Williams offers good advice. In particular, what we must guard against is a behaviorist theory of motivation, as we may call it, according to which motivation is exhausted by what directly issues, or by what is disposed to issue under certain conditions, in observable behavior. Now since everything can be causally connected to anything under certain conditions, it may be hard if not impossible to find counterexamples the behaviorist approach cannot handle. I shall still try momentarily. Yet even if we could place every motivational state in relation to some behavioral disposition, a major complaint remains, concerning how well we would be positioned to understand motivation, exclusively viewing it in relation to cause observable behavior under certain conditions. What is most troubling is the tendency of the behaviorist approach to disregard the inner life of motivation altogether, leaving out entirely the distinctive perspective characteristic of the agent thus motivated. When we are motivated to do something, we usually don't experience that motivation as a simple drive to behave in a certain way, quasi as an internal pull in some direction, but rather experience it as a

rich phenomenon that involves a certain perspective of appraisal on the world and ourselves.

I will return to appraisal and the distinctive perspective inherent in motivational states later when I discuss concern and appeal. For now I would like to challenge the behaviorist approach more directly and turn our attention to a couple of examples of motivational states that appear to bear no immediate or dispositional connection to observant behavior at all, attempting to mount some counterexamples after all. The examples all involve various forms of desire, the quintessentially motivational state, firmly placed on the non-representational and world-to-mind-direction-of-fit side. The scenarios are inspired by Timothy Schroeder's excellent discussion on desire and motivation *Three Faces of Desire*, even though Schroeder adduces the very same examples to support a seemingly opposite conclusion, namely that a purely motivational account of desire is inadequate. The discrepancy does not run very deep, though, since the account of motivation Schroeder is working with is the very same behaviorist account. Applied to desire, his critique takes aim at what he calls the *Standard Theory* according to which "To desire that *P* is to be disposed to bring it about that *P*." (2004:11) His complaint that desires need not be motivational in that sense thus aligns quite congenially with my case for a broader construal of motivation, an account transcending the restriction to dispositions to arouse observable behavior under certain conditions. Anyway, here are the examples.

Schroeder imagines the case of "an ancient Greek mathematician who is uncertain about the value of π but who desires that it not be expressible as a fraction of two natural

numbers.” He notes that “there is nothing our mathematician can do, and nothing he believes he can do, to affect the value of π .” Addressing the artificiality of the case, Schroeder continues: “Desires regarding the necessary facts of the world, such as facts of mathematics, are admittedly rare, but not as scare as might be imagined. ... I might desire that I had never been born, or that my parents had never met, or that, right now, life exists elsewhere in the universe (or, to be more prosaic still, I might desire that my girlfriend, away at a conference, not be having an affair at this very moment). Because of the necessary facts about causation, these are also desires for ends I can do nothing to bring about.” (2004:16) There simply exists no situation or set of circumstances to place our mathematician in where he, pulling off the knowingly impossible, could actually be doing something about the value of π - the example was chosen precisely for this impossibility of effect, documenting an instance of desire that cannot be accommodated in virtue of its disposition to do something under certain circumstances.

This inspires the imagination. I would bet most of us have plenty of desires of that kind. Desires about the past come to mind, that the holocaust did not happen, that the Roman Empire did not collapse, that dinosaurs did not die out, at least not the cuter and more cuddly ones. Yet we know we cannot change the past. Another rich repertoire of knowingly unattainable desires consists in self-regarding desires. I most definitely desire to live longer than I possibly can as the biological creature I am, and also desire to differ biologically, even with regard to features that cannot feasibly be co-designed, to have wings, more powerful organs, additional legs and arms, and most importantly, a larger brain with hitherto unimaginable cognitive powers, and yet to be light and flexible. And then Nomy Arpaly at times desire to be Bertrand Russell. (2006:XXX) Putting together

desires about the past and oneself brings us to another rich domain of desires known to be wanting of all measures to bring about their object: regret. Often what we desire most is that we had not acted as we did. When we regret what we did, it is not just that we believe we acted wrongly or inappropriately, or that we firmly resolve not to do the same in the future. The unmistakable content of our desire *that we had not done what we did* together with our certain knowledge that we cannot change our past, distinguishes regret as precisely the tragic experience it is.

Taking this one step further, it may be questioned, despite widespread acceptance, whether desire, and by extension motivation, bears any essential connection to action, unless the concept of action itself is construed so broadly as to include the entirety of mental activity, including pure motivational mental activity, such as feeling bad about certain things about which one evidently cannot do anything, taking pleasure in imagining oneself to do the impossible, wishing what cannot happen nonetheless would happen, etc. Galen Strawson has produced an intriguing thought experiment designed to fuel the suspicion that there could be desires with no link to action whatsoever. Setting the stage for it, Schroeder writes: "The claim that desires are, necessarily, things that dispose us to action gives rise to other problems as well. Why could there not exist beings quite incapable of action who nonetheless had desires?" (2004:20) Galen Strawson then invites us to imagine creatures of precisely that kind, the Weather Watchers. Endowed with cognitive capacities, enabling them to reliably track the weather, the Weather Watchers remain quite incapable of doing anything about it and yet harbor strong desires that the weather be one way rather than another. Their lot is one of mere spectators, or, to use Galen Strawson's term (1994:254), one of "motionless meteorologists," which does

nothing to detain them from taking an active interest as what the weather is like. Here is Galen Strawson's introduction (1994:251):

The Weather Watchers are a race of sentient, intelligent creatures. They are distributed about the surface of their planet, rooted in the ground, profoundly interested in the local weather. They have sensations, thoughts, emotions, beliefs, desires. They possess a conception of an objective, spatial world. But they are constitutively incapable of any sort of behavior, as this is ordinarily understood. They lack the necessary physiology. Their mental lives have no other-observable effects. They are not even disposed to behave in any way.

After raising the question "Are the Weather Watchers impossible?" Strawson proceeds at great length to make plausible that the answer is no. His case strikes me as intuitively compelling, and his reply to objections of various kinds convincing. Strawson dedicates an entire chapter on the case, aptly titled *The Weather Watchers*, to be consulted in the event doubts remain. Besides his intriguing scenario, there are many more examples that could be provided to challenge the questionable necessity of immediate or dispositional connections between wanting something and doing something to bring it about, and reasons of space permit to consider only one final group. Often it is a core component of our desire for something to happen that we did not have any part in it. We want our children to succeed on their own, a friend to get his deserved recognition without our intervention, our spouse to love us without us prompting that love, dear people to notice when they offended us without us reminding them, receive an invitation to the birthday party without our initiative, etc. Often the nature of what we want is such that the deed would be spoiled if it was us who brought it about, however indirectly.

None of this is to deny the important connection to action many, perhaps most, of our desires exemplify. The examples cited in the last few paragraphs may strike us as unusual, and yet they serve as a cautionary note, reminding us that not everything on the non-representational divide of our psychic life is in the business of making us do something.

5.6. Concern and Appealers

We engage in practical deliberation because we care about things, because we harbor concerns. Concerns represent a major element driving practical deliberation, and the remainder of the chapter is dedicated to describe their nature in greater detail. What I am going to say about concerns may appear stipulative on occasion. That's fine. This is not primarily an exercise in conceptual analysis, but an attempt to delineate and draw attention to a phenomenon of first deliberative importance. I also prefer speaking of concerns, and alternatively of caring, instead of desires, the more customary choice in many practical treatises, not because I deny that concerns are desires, which they obviously are, but because I'd like to single out concerns as a subspecies of desires playing a particularly prominent part in the kind of deliberation I am interested in. As we shall see, concerns present a particularly rich, integrated and temporally extended species of desire, displaying also an interesting combination of specificity, diversity, and particularity. What is it to be concerned, then? First and foremost, being concerned stands opposed to being indifferent; an indifferent creature would find little point in pondering what to do, since nothing it could do would make any difference to the creature. It would

be neat if the English language allowed for a direct antonym for the condition called *being indifferent*; the obvious choice would be *being different*, unfortunately it already has quite a different usage. To be concerned is to be different, then, in precisely this sought after sense. Frankfurt offers the same idea (1982:269).

It is necessarily the case, of course, that a person who cares about a certain thing is not cold-bloodedly indifferent to it. In other words, what happens to the thing must make a difference to a person who cares about it, and the difference it makes must itself be important to him.

When something makes a difference to a person, *it really can make a difference to the person*, in that how well his life goes becomes tied up with how well his concerns pan out. Here is Frankfurt again (1982:260).

A person who cares about something is, as it were, invested in it. He identifies himself with what he cares about in the sense that he makes himself vulnerable to losses and susceptible to benefits depending upon whether what he cares about is diminished or enhanced.

The corresponding worldly counterpart to concerns – the possible or actual state of affairs realizing the content of the concerns – I propose to call *appealers*, an intentionally idiosyncratic terminological choice, primarily designed to underline the contrast with value, or at least with value normatively understood. There is another sense of value appeal aligns with rather nicely, a non-normative sense, namely the sense we have in mind when we speak of the *market-value* of some asset. The market-value of some asset is a direct function of what people are inclined to pay for it, as opposed to exemplifying the normative condition of rendering appropriate or inappropriate what people are inclined to pay. Appeal is like market-value in this regard, pertaining to the

domain of what is actually desired and how much, unlike normative value, pertaining to the domain of the desirable, of what is fit to be desired (incidentally, the notion of *appeal* thus comports well with the notion of *utility* as economists are prone to using it). We are concerned about what appeals to us, and what appeals to us is that which we are concerned about, and thus I explicitly refuse to imbue the notion of appeal with the normative connotations characteristic of value, bearing the altogether different relation to concern, precisely the kind of normative relation I deny, namely that of *meriting* or *warranting* or *rendering appropriate* concern. What appeals to us is not what renders concern *appropriate*, as value would, but simply what *renders* concern, to excuse this instance of bad English in the service of providing the contrast with greater poignancy. People who care about philosophy are people to whom philosophy appeals.

Eventually we might wish to broaden the category of appeal, by loosening its restriction to *actual* concern and by permitting it to include what we would care about under certain conditions, turning it into a dispositional notion. The reason is that the notion of something we would have cared about had it only come to our attention serves an important deliberative purpose, and the term of appeal is in part introduced to serve that purpose. It rings true to say that philosophy was an appealing career choice to me even before I was able to acquire the kind of familiarity with it required for fostering the concerns I harbor for it now. But even under this more extensive and dispositional usage, appeal still lacks the normative connotation of rendering anything appropriate. An applier is simply what would prompt a concern under certain conditions.

Concern and appeal are two sides of the same coin, relating to each other as motivational push and motivational pull, directing attraction and drawing attraction simultaneously. Neither should claim priority. Perhaps one is inclined to think concern comes first and appeal second, with concern having explanatory priority over appeal? This is what several subjectivist treatises suggest, quite shortsightedly, rendering a huge disservice to their own position by making it appear incompatible with some rather evident truth, encouraged perhaps by the ill begotten language of desires and concerns *conferring* significance, value or appeal onto their objects, making the former's priority over the latter almost seem inevitable. When we come to understand a concern, however, we usually must also mention something about its object, something that makes it appealing. More often than not do we find a lot to be said about why we are concerned about what we are concerned about, and in particular, find things to say pertaining to the object of our concern rather than our own mere sentimentality. The most seductive caricature of subjectivism – fostered by a few subjectivists themselves – is that the only thing subjectivists can point out is that we happen to be attracted to certain things and that's that, with no further fact belonging to the object-side that could render the attraction intelligible.

Fortunately, subjectivism is quite capable of withstanding the caricature, becoming apparent as we turn our attention to the prime case of the subjectivist attitude-object relation: taste. Philosophers usually do not shy away from taking a subjectivist approach on taste, or major portions of it, concurring with Don Loeb (2003) in finding comprehensive Gastronomic Realism rather distasteful: *De gustibus non disputandum est*. Indeed, discontentment with subjectivism in the moral, practical, and aesthetic domain is

commonly articulated in terms of taste, the complaint being that the good, rational, or beautiful cannot just be a matter of taste as subjectivism allegedly would have it, presupposing in this very critique that subjectivism must be correct vis-à-vis taste. Center in on taste, then, and in particular on the taste of chocolate, a source of delight the Swiss author enjoys great familiarity with, and suppose there really is no objective realm of chocolaty evaluative facts floating around. Still, when asked as to why one is attracted to one specimen rather than another, the gourmet is capable of responding at great length and in great detail, insisting that his gustatory approach, at least in part, is one of *discernment* instead of *projection*. The nutty taste, its richness, its complexity, its ... There is a world of difference between a fine Swiss piece of a Lind truffle and an American piece of Hershey “chocolate,” manifest even in the raw bite alone, of soft smoothness versus a sandy feel of unabsorbed sugar (to be fair, America has Bissinger).

In sum, I propose to think of concern and appeal more of a unified phenomenon, the concern-appeal complex, involving two sides we may isolate in analysis, where neither side is best seen as dominating the other. The term ‘concern’ can actually be used to stand for either side, as when we can speak of a concern as an attitude or alternatively as an object of our attitude, for instance when we say of something that it is a major concern of ours, the mounting debts of most nation-states, for instance, the concern here being a state of affairs causing us to be concerned. My emphasis of the object-side was not intended to downplay the subject-side. What is appealing is so always relative to specific subjects with specific sentimentalities and specific concerns. My emphasis was driven by the worry that frequently the object-side drops out of the picture entirely, overlooking that only in virtue of specific features can anything be appealing in the first

place. Aspects of the attraction and aspects of the object of attraction both matter explanatorily. The complete picture will be interactive, such that to explain concern and appeal alike we must mention features pertaining to the subject-related and object-related side alike, where variations in the features associated with the object might just as well result in the withdrawal of its appealing character as variations in the features associated with the subjects.

The concern-appeal relation is in this regard similar to a whole host of other familiar relations. That a positively charged field of magnetism dispels a negatively charged field must be explained both in terms of the negative and positive charge of the fields respectively. That we are incapable of digesting little iron balls has just as much to do with the chemical composition and solidity of the iron balls as the structure of our digestive system. That we occasionally experience the unfortunate condition of influenza has as much to do with the structure of the virus as with the structure of the host. Another case in point might be the color/perception pair, or more generally the complex of secondary qualities on the one side and the corresponding receptive capacities on the other, a model famously introduced by John McDowell for thinking about the relationship of value and our sentiments. McDowell's introduces the model to challenge the notion that one or the other must have priority, and in this I wish to follow him. He also takes this one step further, however, speaking of values meriting our sentiments, and here things become more problematic, as we shall see later on.

With appealers, the object-side of concerns, duly noticed, we are in a much better position to capture practical deliberation within my overall account. We can now reap the

profits with regard to two important and interrelated points, one more positive and one more critical, both concerning the reflexive deliberative perspective. The first speaks to the phenomenology of deliberation, reminding us of some crucial observations from the third chapter, namely the fact that when we deliberate, we rarely pay attention to our concerns themselves, which are featuring in the deliberative background as opposed to the deliberative foreground. Since the crucial lesson, I believe, was rather nicely expressed in that chapter, I will rehearse it once more.

Phenomenologically speaking, deliberation engages the world, being outward-directed rather than inward-directed, and in order to accommodate that fact we need to understand how precisely concerns contribute to the deliberative enterprise, and in particular, where and how they enter the deliberative theater. Deliberation engages the world and rarely amounts to a mere exercise in navel-gazing. To be concerned is to have a stake in things, but what we have a stake in are things and not concerns. What we should say is that concerns constitute our deliberative background, and that they only rarely, only under especially peculiar circumstances, occupy our deliberative foreground. (Enoch forthcoming c, Schroeder 2008a & 2008b, Dancy 2000, Blackburn 1998, Pettit and Smith 1990) What recommends certain choices instead of others first and foremost has to do with how things are. This is not to diminish the important role of our concerns, for without them no situation by itself does ever recommend anything. Yet as important the role they play, the role is better seen as that of a background enabler, which is that of a spotlight operator rather than that of a frontal stage actor. We look at our options through our concerns instead of paying them more direct attention. It is in light of our concerns that certain practical options become more attractive than others, where

importantly the light springs from our concerns instead of shining on them, and where what is illuminated in particularly favorable colors are choices rather than concerns. What we have in sight when we deliberate are states of affairs and their appeal on us, not stats of our minds.

This leads to the second point which is directed against a core doctrine in the philosophy of Harry Frankfurt's, namely the alleged centrality of higher-order desires and volitions for enabling reflexivity in concern-related deliberation. In large parts, my discussion on concerns could be regarded as an ongoing friendly conversation with Frankfurt, whose work has taught me to appreciate many of the underlying problems and to whom many of my central observations are owed. The centrality of higher-order desires within his philosophy, however, that is of desires whose object are lower-level desires, has always profoundly puzzled me. For Frankfurt, our ability to form higher-order desires is critical for our ability to engage in reflexive deliberation, our capacity to be deliberators instead of mere wantons, and henceforth desires and concerns without endorsement from higher-order desires don't fully matter. (It is worth mentioning that Frankfurt is not the only proponent of *the hierarchical model*, emphasizing higher-order desires; another is David Lewis, especially in his essay *Dispositional Theories of Value*, in Lewis 2000:68-94) The reason why I bring this up now is that with appealers in sight, we can appreciate how the reflexive stance characteristic of concern-related deliberation can more handily be accommodated with regard to what appeals to us than with regard to higher-order desires. But first, here is Frankfurt (1988:163-164):

There is also another sort of reflexivity or self-consciousness, which seems similarly intelligible as being fundamentally a response to conflict

and risk. It is a salient characteristic of human beings, one which affects our lives in deep and innumerable ways, that we care about what we are. ... We are particularly concerned with our own motives. It matters greatly to us whether the desires by which we are moved to act as we do motivate us because we want them to be effective in moving us or whether they move us regardless of ourselves or even despite ourselves. In the latter cases we are moved to act as we do without wanting wholeheartedly to be motivated as we are. Our hearts are at best divided, and they may even not be in what we are doing at all. ... This means, moreover, that we are to some degree passive with regard to the actions we perform. For in virtue of the fact that we do not unequivocally endorse or support our own motive, it can appropriately be said that what we want – namely, the object of our motivating desire, and the desire itself – is in a certain sense not something we really want. ... It is possible for a human being to be at times, and perhaps even always, indifferent to his own motives – to take no evaluative attitude toward the desires that incline him to act. If there is a conflict between those desires, he does not care which of them proves to be the more effective. In other words, the individual does not participate in the conflict. ... Since he exercises no authority, by the endorsement or concurrence of which certain of his desires might acquire particular legitimacy, or might come to be specially constitutive of himself, the actions engendered by the flow and clash of his feelings and desires are quite wanton.

The passage contains a number of points and moves I find troubling, for instance the artificial and awkward split in personality it suggests between our regular desires and motives on the one side and some overseeing second-order part on the other side, which strikes me as an altogether unhelpful and homuncular model, facing “the problem ... whether we can make sense of these ideas ... of ownership and rejection of a desire,

without appeal to a little person in the head who is looking on at the workings of her desires and giving the nod to some but not others.” (Bratman 2003:221) Related to this is the poignant criticism mounted by Gary Watson (1975) as well as others, regarding the questionable authority and privileged status higher-order volitions allegedly enjoy simply in virtue of their formal position within the overall desire-architecture. Yet all this does not quite get to the much simpler point I wish to stress now, stemming from a puzzlement emerging at an earlier stage still, flowing from my inability to recover within my own deliberation-oriented observations the phenomenon of higher-order desires in the first place, except as a rare and artificial appearance, and quite not as the pervasive phenomenon Frankfurt suggests. Moreover, it seems to be altogether unnecessary, since there is a rather straightforward solution for how to accommodate reflexivity without the recourse to higher-order desires. This would be a welcome result, since, as what is now a philosophical commonplace, appealing to higher-order desires just reintroduces the problem of reflexivity with regard to them, immediately giving rise a question Frankfurt himself recognizes, namely “in what way an individual with second-order desires or volitions may be less wanton with respect to them than a wholly unreflective creature is with respect to its first-order desires.” (Frankfurt 1988:165) The regress his model threatens to get us into, to yet higher and higher order desires, is best avoided right away and not when it has already been set in motion.

Let me begin by trying to make my puzzlement more transparent before turning to the solution. I believe I have a pretty decent grasp of what I want from life, but virtually no grasp of what I want to want from life. Here’s an example. I love skiing, there is no doubt about that in my mind; I do not, however, love camping, to add an explicit contrast

case. Yet when asked whether I also I want to love skiing, rather than want to love camping, I have little clue as what to say. To the question *whether I want to spend time skiing* I am able to answer with an emphatic *yes*; to the question whether I also want to want spend time skiing I am unable to provide an answer one way or the other. Rephrasing the question in terms of what kind of person to be – whether to want to be that person over here full of fondness of skiing and devoid of fondness of camping or whether to want to be that person over there with the reverse fondness-distribution – doesn't help much either. In one interpretation, the question is clear and so is the answer, but that interpretation is rather trivial and quite unlikely what Frankfurt is after: Here I am, with my fondness of skiing, calling to my mind the two scenarios and find that in the second one, where I am the person who does not want to spend time skiing, *I actually don't spend much time skiing*, which is definitely not something I want (i.e. I want *to spend a lot of time skiing*). Assessing this second scenario through the perspective of my extant desires I unsurprisingly chose to reject it. In the end, what I'm inclined to say to all this is that I am just very fond of skiing, that there's a lot I find going for it – the speed, the freedom, the ease and elegance of movement, everything, really – and that's that. I might add that I harbor no regrets having spent time skiing (except from a financial perspective), that I find the activity deeply rewarding, that my high expectations about it are usually met, etc. In short, there is a lot I can say in the way of reflecting upon my relationship to skiing without locating any second-order desires towards my skiing-fondness.

This leads to what seems to me a rather straightforward solution for how to accommodate reflexivity in caring-related deliberation. We are reflexive caring-creatures

(carers?) because we have the capacity, having actual and potential appealers in view, to ask *what to care about*. We let our concerns engage the world with often stunning results. Common language offers a clue: when unsure about whether we want what we think we want, we usually phrase the question thus: *Is this really what I want?* You think *this* is what I want and then, having thought more about the *this*-part, you change your mind: no, *this* is not what I want. Perhaps you noticed some further details that change the overall attractiveness of *this* situation, perhaps you have worked out what *this* involves, perhaps you came to notice expectations raised by *this* not met by reality. Or you think *this* is what I want and then, calling *this* situation to your mind, other concerns of yours speak up, protesting loudly and clearly, repelling you from your initial attraction to *this* (although, as we shall see later, this idea must be treated with great care). To reiterate, *deliberation engages the world, and to be concerned is to have a stake in things, but what we have a stake in are things and not concerns*. Baring especially unusual circumstances, deliberating about concerns is first and foremost deliberation about the relevant object of the concerns, and thus deliberation again and again ends up at the same spot: as a special form of thinking *about the world* (as it is or could be). Contrary to Frankfurt, not only is this form of object-orientation in deliberation compatible with reflexivity, it actually makes better sense of it. It serves to underline how important *appeal* is for understanding practical deliberation.

5.7. The Resilience of Concerns

Concerns are governed by a host of norms specifying when and under what conditions specific concerns may be said appropriate, norms crucial for explaining important elements of the phenomenology of concerns. The subjectivist should not gainsay the infusion of norms into the domain of concerns, for without them, as the following section shall document, he would stay incapable of accommodating crucial features of our concern-structure. At the same time, in thinking about what norms are governing concerns, sufficient attention must be paid to the resilience of concerns making a mockery of overly simplistic treatments. For the purpose of specifying the content of the relevant norms, we can isolate a couple of elements of resilience that ought to be respected, elements providing relevant data-points for future theorizing about concern-governing norms and also serving as cautionary notes against overly simplistic treatments of the norms pertinent to concerns, allowing us to make some reasonable conjectures as where further inquiry into concerns may lead. What our target theory has to accommodate is the specificity, diversity, particularity, and reflective and cross-temporal stability concerns display. These five elements of resilience concerns display are instantaneously recognizable once we attend specific examples, and are all interrelated, especially the first three. I shall discuss the elements of diversity, particularity and reflective and cross-temporal stability first, and turn to the element of specificity later in the context of the discussion on appraisal.

Begin with diversity, or a certain kind of disunity within the plurality of our concerns. In a very broad sense, the relevant concern-governing norms could all be

classified as norms calling for mutual coherence, but this notion needs to be treated with great care. A minimal mark for concern-coherence, let us suppose, is joint satisfiability, the possibility of integrating all relevant concerns into a single life plan, where this principle of agglomeration presumably specifies at most a necessary but not a sufficient condition for concern-coherence. Without further engaging in formal theorizing about what criteria define concern-coherence, we need to turn our attention to the important point that there are going to be limits on the scope of coherence, in any plausible construal, and the following paragraphs are partly intended as a cautionary note against the familiar subjectivist knee-jerk response to cite coherence as *the* criterion for managing concerns.

What we need to come to terms with is that we are susceptible to a wide and diverse range of good things, harboring many concerns pulling in quite different directions, rendering their harmonious agglomeration impossible, yet without this kind of joint non-satisfiability discrediting or undermining the assembly of disjoint concerns, rendering it illegitimate in some sense to maintain them all. English gardens provide a better picture for our overall concern-structure than French gardens. There simply are many good things in life we can be fully appreciative of while also being cognizant of the fundamental impossibility of having them all, and that organizing a life around some concerns may systematically place others beyond that very life's reach. It would be a dubious ideal directing us to force our entire network of concerns into some coherent superstructure, for this would inevitably blind us to aspects we deem important even though we recognize are jointly unattainable. If we were to suppose practical rationality dictates complete coherence in concerns, calling us to move towards maximally coherent

desiderative profiles, considerations of practical rationality would as a result have diminishing capacities to bother us. To excuse a perhaps overly dramatic tone, I worry about totalitarian ideals calling for some sort of concern-cleansing in the name of overall homogeneity.

Examples help to render the phenomenon of concern-diversity instantaneously recognizable. There is something to be said for a life of surprise and spontaneity, and there is something to be said for a life of steadiness and stability, diverging lifestyles expressive of two different temperaments, one resembling more the New England climate and one resembling more the Southern Californian climate (this is a popular example; see Velleman 2009). Now if one is to guide one's life by one ideal or the other, in such a way as to actually enjoy its benefits, one must be prepared to take steps which systematically put out of reach the benefits of the other life. A career promising stability will permit only so much spontaneity. Leaving one's options open for switching occupations at any moment is going to be incompatible with carrying out those occupations that require many years of training and discipline. A social life of constant change is going to be at odds with fostering the kinds of personal loyalties and attachments characteristic of a social life cherishing stability. The ability to relocate and explore new corners of the planet, just like that, places an upper limit on how much at each station one can accumulate which one would, in the transition to the next station, inevitably have to leave behind. It does not require unusual imagination to fill out further details on how deep the divergence in direction runs between a life of stability and a life of spontaneity. The crucial point is this: one can fully embrace one life and forsake the other, assigning greater significance to the benefits the first affords while being prepared to accept the

surrender of the benefits the second affords, without becoming less appreciative of the fact that the chosen life has to leave out some of the things one cares about as well.

The impossibility of agglomerating all of the concerns underlying these two ways of life only illustrates in particularly grand terms what is common from mundane experience as well. One may really want *that* delicious appetizer and *that* delicious desert, knowing together they will exceed the available space, want to spend the upcoming holidays with the extended family and go skiing in a cozy resort, want to write a dissertation on normative reasons and on creativity (or what not), want to live in the warm but boring south and in the cold but beautiful north, or to want a thousand of the other mutually incompatible good things in life. In fact, so familiar is this phenomenon that we are equipped with various psychological mechanisms helping us to cope, most prominently the sour-grape-syndrome: can't have it thus don't really want it. Recent empirical research has validated this phenomenon up to some point, the inclination for the sake of preserving sanity to cease caring for things we just cannot have (Daniel Gilbert 2007). But it is also true that this supplies a healthy remedy only up to a point, and only simpletons would be willing to adjust their appreciative capacities to precisely the circumscribed sphere of what little they can actually get. In reverse, I think there is also a case to be made for maintaining lots of opposing concerns, again in the name of preserving our sanity, by equipping us with a certain degree of concern-flexibility in the face of life's tendency to frustrate our plans, enabling us to devise alternative yet similarly gratifying life-plans. Similarly, maintaining a rich set of concerns brings with it an important side benefit, empowering our capacity of understanding, especially with regard to people who have chosen very different life-paths in very different contexts,

enabling us to recover the point behind their aspirations and pursuits within our own (dormant) concerns.

Notice what is true of incoherent concern-sets is not true of incoherent sets of intentions and plans. With regard to intentions and plans it is plausible to suppose that internal disorder is characteristically problematic, that something inherent to the nature of intention explains why incoherent sets of intentions are ipso facto bad sets of intentions just as something inherent to the nature of belief explains why inconsistent sets of beliefs are ipso facto bad sets of beliefs. Various authors apparently have failed to fully comprehend that there is a relevant difference between beliefs and intentions on the one side and concerns and desires on the other side. Michael Smith (1994:159) writes:

[W]e may properly regard the unity of a set of desires as a virtue; a virtue that in turn makes for the rationality of the set as a whole. For exhibiting unity is partially constitutive of having a systematically justified, and so rationally preferable, set of desires, just as exhibiting unity is partially constitutive of having a systematically justified, and so rationally preferable, set of beliefs.

Granted, disunity, or inconsistency, in sets of beliefs is a sure sign faulty elements have crept in, as in such a case not the entirety of its members can be true; likewise, disunity, or internal mismatch, in sets of plans and intentions is a sure sign that something went wrong, since plans and intentions are in the execution-business, which in turn explains why incoherent and hence non-executable plans and intentions are flawed. In sharp contrast, however, disunity in sets of concerns indicates no such internal flaw, since caring is not in the execution business, or for that matter any business, in the first place. We simply care about what we care about, even if it is impossible to always find

expression of the entirety of our concerns within a single life-plan. Moreover, planning is recognizably secondary to caring, relating to caring more as a servant than as a master, a process somehow informed by our concerns, without, however, this to involve any automatism or immediate translation – or need thereof – from concerns to plans (another indicator of the inadequacies of the simple behavioral story). Planners have to figure out what to do in light of what they care about, a very complicated task indeed, but the standards that apply to that task need not also apply to caring itself. It is entirely reasonable, for instance, to abandon an intention one realizes is unattainable, but not reasonable – or at least not unreasonable not – to abandon a concern one realizes is unattainable; recall Schroeder's mathematician desiring that π not be expressible as a fraction of two natural numbers, or Nomy Arpaly desiring to be Bertrand Russell.

Thus, my inclination to find attractive to spend the evening at home and to spend the same evening at the party can peacefully coexist in a manner in which my intention to spend the evening at home and to spend the same evening at the party cannot just as my belief that I am going to spend the evening at home and that I am going to spend the same evening at the party cannot. It is interesting to note, in this context, that belief and intention are closed under conjunction in a manner in which desire is not. That is, if one believes p and also believes q , one has grounds to believe $p \ \& \ q$ (or ought to, or is rationally required to, read wide scope, cf. Broome 1999), just as if one intends p and also intends q , one has grounds to intend $p \ \& \ q$ (though I admit I'm less confident about the intention case than the belief case). In contrast, if one desires p and also desires q , one need not have grounds to desire $p \ \& \ q$. For instance, I may want Peter to join the party, and also may want Lisa to join the party, but may definitely not want *Peter and Lisa* to

join the party, since they always quarrel and would ruin the party. Thus, the contradiction between different conflicting beliefs and intention creeps into the content of some particular belief, yielding a belief with a contradictory content, while the incoherence between different desires does not creep into the content of some particular desire, yielding no desire with incoherent content. As we shall see momentarily, there is another respect in which it is crucial to pay attention to the plan-concern dichotomy.

There's a prominent philosophical stance which at first acknowledges the inalterable concern-diversity as part of our lives, but then goes on to deplore that diversity as non-ideal, as something we regrettably have to live with but which at the same we recognize as suboptimal. I find such idealizations entirely unappealing and beside the point, to the extent to which I even understand them. What is the corresponding idealization where the conflict between a concern for a life of stability and a concern for a life of spontaneity is removed, and why would *that* scenario be relevant for thinking about a situation where the conflict is not removed but very much alive indeed? Here I am, I say, with my not fully coherent and diverse concern-sensibilities, facing the task of figuring out what to make of such diversity. What are ideal worlds to me where harmony is established by silencing some of what I care about? This is a sentiment well expressed by Isaiah Berlin (*The Pursuit of an Ideal*, 2000:11), speaking of values instead of concerns:

These collisions of value are of the essence of what they are and what we are. If we are told that these contradictions will be solved in some perfect world in which all good things can be harmonized in principle, then we must answer, to those who say this, that the meanings they attach to the

names which for us denote the conflicting values are not ours. We must say that the world in which what we see as incompatible values are not in conflict is a world altogether beyond our ken; that principles which are harmonized in this other world are not the principles with which, in our daily lives, we are acquainted; if they are transformed, it is into conceptions not known to us on earth. But it is on earth that we live, and it is here that we must believe and act.

In closing this cautionary note on the scope of coherence as an ideal for concern-sets, designed to encourage us to embrace the concern-diversity found in our hearts, two points bear mentioning. First, the kinds of reservations I voiced about the high esteem of desiderative *coherence* as a virtue held by Smith and others does not entail esteeming desiderative *incoherence* as a virtue; I merely believe it is a deep fact of human psychology reflecting our attraction to the many diverse good things in life. Secondly, I expressively do not deny that coherence ever has an important role to play. As we shall see, within particular networks of interrelated concerns, concern-clusters, coherence becomes quite important. With this, turn to another data-point concern-related theorizing ought to respect.

This second aspect is particularity, or the characteristic context sensitivity concerns display. (Dancy 2004a; Kagan 1988) We best appreciate what this second aspect amounts to by turning our attention to what it stands opposed to, namely a seductive yet also quite simplistic analysis of concerns which breaks them down into some assembly of smaller sub-concerns, taking the overall concern to be the sum of its parts. Call this the Lego-brick model of concerns. Here's a big concern, the overall attraction, and here are its parts, its sub-attractions. We explain the big concern by

isolating its various sub-attractions, and then by adding up these sub-attractions we determine the overall attractiveness-score. Part of the appeal of the Lego-brick model is that it promises some neat manageability of concerns, by proposing a certain isolation-procedure: explain overall attraction by isolating individual attractive parts. This would make for a simpler theory, supplying some neat generalizations: wherever one can identify feature such-and-such one will find some attraction towards it as well (provided certain conditions obtain, such as awareness of the relevant feature). As a theoretical ambition, however, simplicity is a virtue only if the target phenomenon is simplistic too and is bound to lead to nothing but distortions if not.

The primary problem with the Lego-brick approach does not necessarily lie in its attempt to understand concerns by breaking them down into various components, as concerns obviously contain components. Rather, the problem lies in the approach's failure to acknowledge the holistic interaction between the various components that are part of the relevant concerns, a kind of interaction that does not permit additive functions from little attractions to big attractions. To the extent to which we are able to identify what particular aspects we are concerned about when we are concerned about something, the capacity of that aspect to arouse our concerns must be understood holistically and within its specific context. The same feature arousing a concern here need not also arouse a concern there, even when we are dealing with the same kind of concern. Concern-invitation is not a feature that could be neatly pinned down to specific bits and pieces, where in turn these bits and pieces could be invariantly paired up with the capacity to arouse concerns across all contexts.

As before, the point is best illustrated by attending to examples, and also as before, gastronomical examples are especially helpful. Consider the enjoyment of a fine piece of chocolate, a Bissinger or Lind truffle. The Lego-brick model suggests that our enjoyment of the truffle is to be explained by some combination of our enjoyment of its various individual features, such as its sweetness and smoothness and so on. It is our basic attraction to sweetness which in part explains our overall attraction to the chocolate piece, according to this model. Yet as a moment's reflection makes crystal clear, this can't be quite right. The delightful Lind experience cannot simply be some felt summation of its attractive parts, but at the very least must consist in their very specific combination. For one thing, it is simply not true that there are these simple taste-features exhibiting some rigid attraction on us, not even sweetness, which would be rather disgusting if part of the tortilla chip or fillet mignon experience. Thus, when we seek to adequately describe the pleasant experience of devouring a fine piece of chocolate, we cannot capture that by merely describing what we like in it, providing some laundry list of likable features, while ignoring the way they combine to the totality of the experience. As every chef knows, professional or hobby, preparing a *tasty dish* is not just a process whereby a bunch of *tasty ingredients* are thrown together. The setting or context of the individual ingredients matters just as much as the ingredients themselves, affecting each other in unpredictable ways, and thus there is no straightforward, let alone additive, function from *tastiness in ingredients* to *tastiness in the resulting dish*.

The same lesson can be applied to concerns more generally, which, to use a popular philosophical term, form *organic unities*. To highlight how the presence of a feature can be central to a first concern while its absence be central to a second concern,

or similarly, how one and the same feature can prompt a concern with positive valence in a first context while a concern with negative valence in a second context, consider a few more examples. I cherish the quiet and solitude of a hiking experience in the mountains, in no minor parts recompensing for the discomfort and strain hiking is bound to involve, yet imagining the same sort of experience of quiet and solitude on Times Square in New York City would be horrifying, conjuring up images from bad Zombie movies. Quiet is a feature arousing a concern in one context (mountains) and repelling a concern in another context (New York City). Similarly, it is perfectly intelligible why someone might be contemplating to spend a lot of money to travel to the Alps in order to have a white and cold Christmas while also contemplating to spend a lot of money to travel to Southern Florida in order not to have a white and cold Christmas, both attractions simultaneously holding sway in his imagination. To accuse this man of confusion, pressing him to make up his mind as to whether or not he likes white and cold Christmases, would precisely betray a failure to understand the particularity of concerns. Or consider a literary example. I greatly enjoy the subtle and laborious style in the writing of Thomas Mann, brimming with self-irony in his refusal to get to the point, but despise the same style in newspaper articles. In similar fashion, I admire the long well-crafted sentences in the writing of John Stuart Mill and equally admire the absence of long well-crafted sentences in the writing of Derek Parfit. Or consider a technical example. One might love a vintage Cadillac for all the little rattles and noises it makes and love a brand new Cadillac for all the rattles and noises it does not make. In sum, context matters a great deal, and the vigor with which context has recently entered the philosophical imagination might be considered *the* philosophical event of the last decade or so. Without the space or ability to

document this further, I have a hunch that there are connections between the various domains in which context-sensitivity has finally been recognized to reign, suspecting, in particular, that the holistic nature of our minds, noted by recent cognitive science, has something to do with the holistic nature of our concern-structures, noted here as well as by recent moral psychology.

This brings us to the two remaining features concerns display in their capacity to anchor deliberation, two more respects in which concerns turn out rather stubborn creatures, namely their reflective and cross-temporal stability, or, put alternatively, their synchronic and diachronic stability. Exhibiting a certain degree of stability is essential to the capacity of concerns to instantiate the resting-points for deliberation. Consider first synchronic stability, which is the capacity of mental items to resist change in the face of revisions occurring at other corners of the overall cognitive-conative system, a form of cognitive resilience or even encapsulation. One important respect in which concerns display synchronic stability is informational: Concerns need to survive the confrontation with a reasonably wide range of factual representations, rather than be founded on evidently flawed information, since concerns which immediately falter after their host has learned of some obvious factual error would provide a fragile foundation for deliberation indeed. Concerns must be compatible with a reasonably clear-eyed perspective on how things are. This analysis partly concurs with the insistence of other authors to place only *well-considered* desires, as opposed to just any desires, at the deliberative center, desires the individual actually holds and has also reflected upon, desires which have withstood the test of reflection and are not experienced as internally flawed or alien or inauthentic. Now I agree this kind of reflective element matters sometimes, but disagree it always or

even usually matters, suggesting that we ought to regard the need for reflection more as special case than the rule. What is crucial for synchronic stability is that concerns do not waver in the face of representational revisions. At times this may require the further condition that they are reflectively endorsed, as a sort of extra safeguard against the threat of reflective destabilization, yet at other times, especially when the relevant concern is not likely to be challenged or subjected to reflective pressure, no such reflexive examination or endorsement is required to secure its stability.

Moreover, informational resilience is usually secured rather handily, at least for our general concerns, which appear to be rather fact-insensitive creatures, immunizing them against factual destabilization from the very outset. This is not intended as a universal claim, for on occasion general concerns are factually conditioned, and furthermore, as we shall see momentarily, particular concern-occasions are usually fact sensitive, but I still believe factual information bears little to no relevance for most general concerns. The point is best appreciated when we once more pay attention to the distinction between concerns and other goal-directed states such as aims, goals and plans which in stark contrast to concerns are heavily fact-sensitive and fact-conditioned. Aims, goals and plans enter the deliberative scene at a different stage than concerns, namely the later stage of planning and decision-making, and which have to be adopted in response to our representations of the world and which, consequently, become immediately subjected to factual scrutiny. When setting goals, we do so in the belief that by achieving the relevant goals, in this world, we are going to realize something of importance to us. Whether the goal meets this expectation depends on the world, and the goal is to be adjusted in case it fails to meet that expectation. In contrast, what is important to us, what

we are concerned about, is not something we set or adopt against a factual background. We don't set concerns but just have concerns and have them for their own sake. We usually find no difficulties in sorting out what matters to us from what we only aim at in order to get something that matters to us. Aims and goals and plans usually bear what is called an instrumental relationship (or some other form of dependence-relationship) to something else, other goals or eventually other concerns; with regard to aims and goals and plans it should always be possible to identify an informative component to complete the sentence "I pursue aim X *in order to, for the sake of* ...". But concerns are never instrumental, but intrinsic, and this usually deprives factual information of relevance from the very start.

General concerns, therefore, are usually fact-insensitive. However, it must be acknowledged, even after paying due attention to the concern/aim dichotomy, that factual information bears considerable relevance for particular concerns, particular *concern-occasions* as I wish to call them. Consider an example. Most people profoundly care for enjoying meaningful personal relationships, to spend time together with people they like and love, and this general concern of theirs, albeit itself a contingent fact, strikingly is not itself contingent upon fact. What is contingent upon fact is whether and with whom and in what manner people are going to enjoy personal relationships – how, in other words, the concern plays itself out, how the *general concern* translates itself into particular *concern-occasions*. The specific practical question, whether to foster a relationship with *Lisa*, is fact-laden; the general practical question, whether personal relationships matter to us is not. In sum, we have *general and fact-insensitive concerns* that express themselves in concrete scenarios, engendering *concrete and fact-sensitive concern-occasions*, as in

our case we have a general concern for meaningful relationships that expresses itself in the concrete concern-occasion of aspiring to foster a relationship with Lisa. Notice concern-occasions are clearly not the same as aims, goals and plans; I intend to write Lisa a card *in order to* make her happy, and the writing is an aim, a chore, something I certainly would not pursue for its own sake. In contrast, I foster the relationship with Lisa *as a way of* living out (but not in order to fulfill) my concern for enjoying meaningful relationships, fostering the relationship with Lisa for its own sake, providing an occasion for the general concern, or, to use a term of Henry Richardson's (1994), providing a *specification* of the general concern.

Yet might not the fact-sensitivity of concern-occasions infect general concerns with fact-sensitivity as well? After all, there is no way of satisfying general concerns other than satisfying particular concern-occasions. In this vein one may wonder: should the rest of mankind turn out to be sense-less automata, or holograms, or complete grouches unworthy of fostering relationships with, our general concern for personal relationships would remove itself, wanting of even a single concern-occasion. Not quite. In such unfortunate circumstance the concern would merely reveal itself unrealizable in this world. Or perhaps one might think if we found out that we only cared for relationships because genes encoding this concern out-reproduced genes encoding opposite concerns, or if we found out the concern was implanted by an evil demon, then our concern would remove itself? Not quite again. In such an awkward scenario the world would merely have a funny history, including our concerns, but unless our concerns have an unusually awkward self-referential content, namely *not to be the product of funny histories*, the concern would still stand. Moreover, it is also not true that

there is a something else, say pleasure, personal relationships happen to afford and in virtue of which we care for them, even though personal relationships of course do afford many other things we also care for, including pleasure. Concerns, then, exhibit synchronic stability in virtue of their capacity to withstand informational updates, among other things, a capacity coming free of charge for most (general) concerns.

Turn then to diachronic stability, the characteristic persistence through time concerns display. Frankfurt once more provides the key insight (1982: 260-261):

[C]aring about something is not to be confused with liking it or with wanting it; nor is it the same as thinking that what is cared about has value of some kind, or that it is desirable. It is especially to be noted that these attitudes and beliefs differ significantly from caring in their temporal characteristics. The outlook of a person who cares about something is inherently prospective; that is, he necessarily considers himself as having a future. On the other hand, it is possible for a creature to have desires and beliefs without taking any account at all of the fact that he may continue to exist. Desires and beliefs can occur in a life which consists merely of a succession of separate moments, none of which the subject recognizes – either when it occurs or in anticipation or in memory – as an element integrated with others in his own continuing history. ... The moments in the life of a person who cares about something, however, are not merely linked inherently by formal relations of sequentiality. The person necessarily binds them together, and in the nature of the case also construes them as being bound together, in richer ways. This both entails and is entailed by his own continuing concern with what he does with himself and with what goes on in his life.

Concerns endure, and thereby provide the glue linking together the moment of thinking about what to do and the usually significant later moment of doing it. Enduring

concerns supply a perspective past planners and future executioners share, rendering their cooperation possible and intelligible, and without which planning agency would crumble into a heap of disjoint time-slices. To appreciate the centrality of enduring concerns for deliberation, I would like to revisit and then build upon an observation from the previous chapter, where I discussed *the constitutive conservative bias in diachronic reasoning and planning*, henceforth *the bias*. The observation I am alluding to is that the very sensibility of forming plans crucially depends upon an agent's anticipated preparedness to follow through past plans in the future; which depends, in other words, on his anticipated preparedness to feel rational pressure in the future towards taking seriously, in a committal fashion, intentions, policies and plans formed in the past. Without that assurance the agent cannot even sensibly begin to form plans in the first place. His present formation of a forward-looking intention is possible only under the anticipation of a backward-looking future self, and in particular, looking back not in a mere act of curiosity but in an attitude of upheld commitment.

To this I would like to add the further observation that only with enduring concerns do we get the right sort of context within which we can situate the conservative bias in diachronic reasoning and planning. To be clear, the further observation I wish to support is not that enduring concerns *ground* or *entail* the conservative bias, for the norms underwriting the bias supply an extra and independent (constitutive) prerequisite to the possibility of cross-temporal deliberation, a norm operating, moreover, at the planning level and not the caring level. It is thus entirely possible that a want-to-be deliberator has enduring concerns without the bias, and consequently, to the extent to which the constitutive claim of the previous chapter proves correct, it is entirely possible

that he might still lack the capacity of long-term deliberation. This could be so because an unbiased deliberator, having a number of equally weighty concerns, and having previously decided to pursue a project specifically tailored to one concern rather than another, still has no assurance that his decision is going to carry over to later moments rather than be reversed, just like that, switching to a project tailored to the other concern instead. All the same, the observation I wish to add is that the bias is intelligibly sustainable only within the context of enduring concerns.

To see this, let us attempt to construct a concrete example of cross-temporal deliberation with bias yet without enduring concerns. Although I cannot establish this conclusively, there is a strong case that such an attempt must fail. As a terminological matter, let us say of a decision that was reached on the basis of or which is motivated by a desire that the decision *expresses* the desire. In the example I wish to construct we encounter an agent who has time-sliced desires but no enduring concerns, enjoying reliable knowledge about desires of the agent's other time-slices. Consider Susan, who, at the beginning of the week, feels very energetic, taking a positive outlook on work, yet who, at the end of the week, feels very lethargic, taking a negative outlook on work. These attitudes are reflected in Susan's desires, being very desirous of working until the end of the week on Mondays but not at all on Fridays. And that's all, let us suppose, explicitly disregarding other desires Susan might have. Now as it happens Susan has to decide, today, on Monday, whether to take off work on Friday; her boss offered her that options but would like to know now in order to make provisions for the rest of the week. It seems clear the only desire-expressing decision she could reach today is to decline the offer and to plan to go to work on Friday. Her present desire to work until the end of the

week is presently the only desire in play her decision could express, and though she knows on Friday she is going to desire not to go to work, this future desire of hers is not presently had but only foreseen. What decisions express, however, are actual desires and not foreseen desires. She knows, of course, that on Friday she is going to regret that decision, but that's Susan on Friday, not Susan today, who desires to work until the end of the week.

When Friday comes along, Susan faces another decision, this time whether to go to work, whether to follow through her previous decision from Monday. The only desire-expressing decision she can reach now is not to go to work; there simply is no desire she currently holds that would correspond to her previous decision to work until the end of the week. Of course, in a plausible real world scenario we would suspect there to be other considerations with relevance, such as not to disappoint her boss who is expecting her at work, but for now we are disregarding them. Now the task is to render intelligible how Susan could still be under the influence of the bias, demanding her to give some weight to her previous decision simply because it was her previous decision. The problem is that there seems to be no rationale rendering intelligible why she should be under this bias, which would seem like a glitch, a case of rule-worship or decision-fetishism.

Now there appears to be a rather straightforward solution to all this, namely to add a desire, or a principle calling for a desire, that one's future desires be satisfied, a future desire-regarding desire. (Alternatively, the solution could consist in adding a backward looking desire, or a principle calling for such a desire, that one's past desires be satisfied, a past-desire-regarding desire. What I say vis-à-vis future-desire-regarding

desires can be applied, *mutatis mutandis*, to this alternative solution). What Susan supposedly needs, on Monday, is a desire that her desire on Friday, foreseen but not yet had, be satisfied. The problem with this alleged solution is that it is difficult to sensibly situate such an additional future-oriented desire within her present desires, and that the principle calling for the inclusion of future-desire-oriented desires is quite implausible. Adding the relevant future desire-regarding desire creates a rather perverse tension within Susan's desiderative profile. The tension is not merely between competing desires, which I argued earlier ought to be accepted, but is of a very specific kind. When there is a direct tension between a present and held desire and a foreseen and not-held desire, there is, from the present standpoint, nothing to be said in favor of the *object* of the future desire (presently there is no attraction whatsoever towards that object). Since deliberation is object-oriented, a positive attitude in connection with a desire must trace, in some sense, a positive attitude towards its object. Yet in Susan's case, on Mondays, there is no positive attitude whatsoever that could be recovered in her present desiderative profile towards the object of the foreseen but not had desire, and henceforth the only rationale left for the disposition towards taking steps to satisfy that future desire could be *that a desire be satisfied*, quite independently of some corresponding appreciation of its object, and that, I believe, is a perverse attitude.

This also explains why we should reject the principle calling for the inclusion of future-desire-oriented desires. To bring this out most forcefully, consider some especially drastic cases where one has foreknowledge of insane future desires: that one's present (later past) projects fail, that one's family and friends be hurt, that the corner of the world one is enjoying to life in be destroyed, or a myriad of other self-and other destructive

desires. The sensible reaction to such foreknowledge of crazy future desires is not to form a present desire that they be satisfied, and as consequence, quasi as a form of perverse self-directed sabotage, start moderating ones efforts towards the realization of one's present projects, or withdrawing from family and friends so as to make harming them in the future psychologically easier, or polluting the sites which one presently enjoys lingering at as to make them inhabitable later. The sensible reaction is to seek help, to safeguard against ones yet unexplained future craziness, doing what it takes to prevent such spiral into psychological disaster (perhaps a product of some "disillusioned" rage against oneself and the world).

This is indicative of a larger pattern of counterexamples. Here is another one, suggested by Trent Dougherty in conversation: Suppose Trent knows that he is going to be corrupted by the security of tenure and will no longer desire to be highly productive after he has secured tenure. The future-desire-oriented principle directs Trent to have a desire to facilitate his future desire, hence to facilitate future laziness, yet *clearly* Trent ought not to have such a desire. (Parfit's Russian Nobleman is of a similar kind.) The diagnosis for why the principle is implausible is clear: When an agent takes an opposite view on the attractiveness of things from that incorporated in his future desires, when he finds himself in substantial disagreement with his future self (*to be of two minds cross-temporally*, as Trent nicely put it), to demand of him to take steps towards furthering his future desires simply because they happen to be his future desires represents a quite dubious demand.

I am aware the case is complicated, and more support would be needed to adequately defend the claim that cross-temporal deliberation *cannot* proceed solely on the basis of time-sliced desires. In particular, I cannot here decisively rule out that a more sophisticated time-sliced-desire version might not do the trick. What should be apparent by now, however, is what *could* rather naturally remedy the present predicament: a common desiderative perspective past and future time-slices are capable of sharing, which past decisions and present executions could simultaneously express. Enduring concerns supply precisely this common perspective, and thus supply the context within which the conservative bias can be intelligibly sustained. A person with concerns conceives of them, and thus to some extent of himself, as persisting through time, taking a *prospective outlook*, and when he reaches a decision on the basis of his concerns he can take some assurance that his future self is still going to recover in his (future) desiderative profile something motivating the original decision in the first place, something setting a context which renders intelligible why to stand by that decision.

Concerns, in sum, are unruly fellows, in virtue of the features of specificity, diversity, particularity, and reflective and cross-temporal stability they display. In order to capture appraisal, we need to situate concerns within larger norm-governed patterns of concerns, without violating, however, the constraints set by the five features. Specificity rules out that concerns be subjected to one-size-fits all norms; diversity adds specification to this by ruling out an all-encompassing norm of coherence applying to the entirety of our concerns; particularity suggests different norms apply to different cases; stability renders the mutual concern-independence even firmer by encapsulating concerns from cognitive revisions made elsewhere in the cognitive system. Adequate treatments of

concerns, as well as of the relevant concern-governing norms, must accommodate all five features, placing us in a position to make some reasonable conjectures about what a successful concern-related theory would look like: Concerns come organized into larger concern-clusters or concern-families, forming interrelated networks and patterns, yet at the same time the norms applying to these patterns cannot be all-encompassing but must be limited in scope, and moreover, allow for considerable independence between different clusters.

5.8. Appraisal – The Problem

Let's return to the concern-appeal relationship. Concerns and appealers are, as I argued above, intimately tied up with each other, and we need to take both into view to fully understand either component. Concerns and appealers are mutually and counterfactually connected – changes in one component will usually result in changes in the other component. Moreover, it usually should be possible to be more precise in explaining particular concern-appeal relationships, and specify what combination of features borne by the relevant applier is causally responsible for arousing the relevant corresponding concern. In such instances we can point towards certain features due to which we are concerned about what we are concerned about. The question we now face is whether this sort of counterfactual relationship is all there is to the concern-appeal relationship. There are grounds to suspect that the answer is no, and furthermore, that close attention to how concern and appeal relate in this further sense threatens to reintroduce precisely the kind of normative element I was previously trying to disown, in

a manner that is more, which makes it quite hard for me to ignore, since I intend to place concern at the center of my *an*ormative story replacing the normative story: namely by finding the normative element to reemerge as an incremental part of the very phenomenology of concern itself.

What is that phenomenology? To answer the question, we must again take as our point of departure the concern-appeal relationship, and in particular, the attitude inherent in concern towards its worldly counterpart, appeal, only this time we must look more closely. Let us speak of the *concerned attitude* or *C-attitude* as the inherent attitude of concern towards the relevant object, appeal. The difficulty is not how the C-attitude and its object could initially become connected. Concerns have representational contents – one is always concerned about something or cares about something, that such-and-such be (or not be) the case, something involving a specific proposition, and thus what state of affairs (appealer) relates to what concern is determined by the content of the relevant concern. Furthermore, the difficulty is not to understand how the C-attitude could be responsive to features borne by the appealers in one important respect, given the counterfactual relation between concern and appeal just outlined. Rather, the difficulty stems from an additional sense of responsiveness that is in play, incremental to concern yet entirely unaddressed thus far. In particular, it appears as though the attitude's relationship towards the content is not merely normatively accidental, even if counterfactually stable, but rather exhibits some specific form of *evaluative receptivity* towards its content. Concern as a whole seems to incorporate an evaluative perceptive element, a particular evaluative perspective on its objects, a particular construal of it, a *charged* form of *seeing-it-as*. Borrowing a term used extensively in the literature on

emotions, I propose to call this element *appraisal*. Concerns contain an appraisal of certain actual or possible state of affairs, or, to use Bennett Helm's useful term, construes the state of affairs as having a certain *import*.

To bring out appraisal more vividly, let me for a moment endow concerns with a perspective, resorting to some hopefully innocent form of anthropomorphism in the service of exposition. Phenomenologically speaking, and from the perspective of the C-attitude, there often seems to be, in the relevant object, an element of invitation *drawing in* the attitude. The attitude does not present itself as a mere blind reaction aroused by certain aspects of its object, which is precisely what we would have to say should the concern-appeal relationship be exhausted by some form of counterfactually stabilized relationship between the C-attitude its object, rendering the relationship between the C-attitude and its object fundamentally accidental. It just so happened, we would then have to diagnose, that given the way the overall system was set up, certain states of affairs with certain characteristics tended to arouse certain concerned attitudes. It would be true to say of the concerned attitude that it was responsive to the relevant aspects of its object *in a triggering sense*. The problem with this is not that it is incorrect as far as it goes, but that it does not go far enough, in particular the problem is that this triggering-sense fails to do justice to the internal viewpoint of the attitude. From that viewpoint, the C-attitude presents itself not just as a reaction to the presence of certain aspects in its object, but as a receptivity to their *attractiveness*, which in turn explains why it might be possible to construe the attitude as more than a blind reaction and instead as a receptive response, as something occasionally *called-for by the attractiveness of the relevant feature*, as

something that can be *appropriate or inappropriate* depending on whether the feature is in fact attractive.

Or so it seems. Phenomenological observations are prone to meet phenomenological counter-observations, and whether an element of receptivity is part of the very phenomenology of concern can be questioned. Hume appears to question it when he writes, in relation to virtues and vices, yet certainly intending his claim to generalize (2000:301):

Take any action allowed to be vicious ... [e]xamine it in all lights and see if you can find that matter of fact . . . which you call vice. . . . The vice entirely escapes you, as long as you consider the object. You can never find it until you turn your reflection into your own breast, and find a sentiment of disapprobation, which arises in you, toward that action.

These are frequently cited words of Hume's, appearing for instance in an article by D'Arms and Jacobson's dealing with the topic of sensibility theories. After providing another citation of Hume's which seems to advance a similar view with regard to beauty – roughly the view, in its common expression, *that beauty is in the eye of the beholder* – D'Arms and Jacobson also give voice to a strong reaction by a contemporary commentator, David McNaughton, who writes: “Hume's confident assertions about the unobservability of beauty are breathtakingly counterintuitive. We see the beauty of a sunset; we hear the melodiousness of a tune; we taste and smell the delicate nuances of a vintage wine. Hume's denial that we can detect beauty by the senses flies in the face of common experience” (1988, p. 55). I am inclined to agree, and so are D'Arms and Jacobson: “Understood as a phenomenological claim, this seems obviously correct – so

obviously that one should doubt whether Hume meant to be denying it. Surely, when we find something beautiful, delicious, or even virtuous, we experience this as a matter of sensitivity to the observed object: the sunset, the wine, the person.” Phenomenological observations are difficult to settle; beyond reporting my readiness to concur with McNaughton’s analysis, I have little to add for the sake of proving the point, and thus one may read the ensuing sections to address an audience sharing McNaughton’s, D’Arms’ and Jacobson’s and my sentiment. The relevant question we need to turn to, together with D’Arms and Jacobson, is this: *But what kind of sensitivity is this?* (Copp 2006:187)

The question leads us to the familiar theoretical choice between detection and projection, between thinking of the relevant feature towards which concerns display evaluative receptivity as being or not being some product of the concerns themselves, or, to use the framework from above, the question is: does the C-attitude toward its object obtain because it has a characteristic feature, of possibly a second-order sort, *attractiveness*, attracting the attitude, or does the object have the characteristic feature of attractiveness simply because there obtains some attraction towards it, a C-attitude? This choice motivates the two standard solutions for handling appraisal. According to a first standard theory, the attitude is conative in nature, for instance some disposition towards furthering (or hindering) the realization of the relevant state of affairs. According to a second standard theory, the attitude is cognitive in nature, for instance some evaluative belief maintaining that the corresponding state of affairs is good, choiceworthy, or supportive of reasons for taking measures to bring it about. Neither theory seems satisfactory. My overall approach is quite clearly inhospitable to the cognitive theory, resting its allegiance with the conative theory, and I am confident further elaboration of

the conative theory promises to bring us reasonably close to capturing appraisal. But first consider why both theories are plagued by problems.

The first account construes the C-attitude as conative in character, some disposition towards seeing to it that the content be realized. The section on motivation already exposed some cracks in the conative account by presenting cases of desire lacking dispositional action-potential altogether. The counterexamples speak for themselves, but I take them to be indicative of a deeper problem still. The problem is precisely the inability of the conative approach to capture the distinctive element of receptivity characterizing appraisal, still construing content and attitude as disjoint components, even though they might exhibit various stable counterfactual relationships with regard to each other. And thus while the conative account can explain how we could come to experience the relevant content in a particularly charged manner, in the familiar sense of the attitude projecting its spin onto its object, it becomes very difficult to see how it could also explain how the attitude is responsive to the experienced spin of its object. In any case, to see how disjoint attitude and content remain in this account, consider an example: Harboring a strong concern for exploring the more attractive corners of this planet, in particular, caring for traversing the Californian Wine country in late summer. Applying the conation approach, here's roughly the analysis of the concern we would get: '[Exploring the Californian Wine country in late summer]: [let's go for it!]' As is readily apparent, this analysis must construe the relation between attitude and content as ultimately accidental, relinquishing all resources to countenance any deeper connection between the two, denying the attitude [*let's go for it!*] to exhibit precisely the element of appraisal, the distinctively charged responsiveness towards the invitingness of

the state of affairs [*Exploring the Californian Wine country in late summer*]. It just so happened, the analysis must suppose, that the subject is so structured as to adopt this attitude instead of the opposite attitude towards it, being drawn to beauty instead of being drawn to Texas' plentiful oil refineries. That the subject's attitude is counterfactually linked to the Californian Wine country instead of Texas' oil refineries quite obviously does very little to help the situation – it doesn't render the concern-appeal relationship any less accidental in the relevant sense. Thus, "the subjectivist model," writes Charles Taylor, "is false to the most salient features of our moral phenomenology. We sense in the very experience of being moved by some higher goods that we are moved by what is good in it rather than it is valuable because of our reaction." (1989:74)

The second account construes the C-attitude as cognitive in character, some form of evaluative judgment affirming the worth and value of the object of the attitude. Some authors have introduced the notion of the "guise of the good" to express this idea. The view is that when we are attracted to or concerned about certain features, we experience those features *under the guise of the good*. This account seems superior to the previous account at first glance. Caring for [traversing the Californian Wine country in late summer] involves seeing something in the activity, and the cognitive account appears well positioned to capture that element. Appraisal appears to be a mental state very closely related to judgment, and since the cognitive account places judgment at the center of concern, it appears well suited to handle appraisal.

Unfortunately the cognitive approach is plagued by problems no less severe, which become apparent when we notice how evaluative judgments are neither sufficient

nor necessary for concern. First insufficiency: One can believe certain activities to be good and worthwhile yet not care for them in the slightest. One can come to truly regard haughty aesthetic exposure as valuable, and yet not care at all for sitting through never ending operas and for wandering through massive museums. One can believe one would be a better person if more sensitive to the needs of distant people but simply not care for being a better person. Or, one can judge some first activity to be better than a second activity, but care more for the second one than the first one, asserting to the greater worth of the discussion group on marital conversation skills compared to the poker round with one buddies, for instance, yet find his concerns unambiguously drawn to the round. Or, we can conceive, as a very general sort of counterexample, of creatures who are apt value judges but don't care for anything. They, The Value Watchers, detect what is the case value wise, similar to Galen Strawson's Weather Watchers watching the weather, but, dissimilar to the Weather Watchers, they don't care a bit. The Value Watchers figure as reliable but disinterested and emotionally detached accountants of value. Perhaps examples like these reflect some rational conflict at some level within their subjects. But even if this was so, these examples provide ample documentation that evaluative judgment does not equal concern. The cognitive account is not supposed to be just an account of the rationality of irrationality of concern, but of concern, period.

Evaluative judgments are not necessary for concern either. One can care for something without holding any evaluative belief about it. This is most obvious with regard to taste preferences. I prefer (care more for) white wine over red wine, but do not believe it to be better in any sense except serving my preference better. And to the extent to which I do hold evaluative beliefs about wine they entirely are an upshots of my

preferences (or concerns), and thus cannot represent something my preferences are responsive of in the first place. Taking this one step further, one can even care for things one believes to be bad. One may care for playing gory computer games while feeling this to be an entirely worthless activity. Or, we can conceive, as a very general sort of counterexample, of creatures who have a value-judgment deficit, unable, for whatever reason, to form value-judgments. Still, they may very well be deeply concerned about how their lives go, that their projects succeed, that their favorite sport teams win, etc.

On a whole, the cognitive account does not appear superior after all. Judgment alone cannot capture the kind of appraisal inherent in concern. Moreover, the difficulties with the cognitive approach reversely mirror the difficulties with the conative approach. There the attitude was construed too hot and too indiscriminate; here the attitude is construed too cold and too discriminate. And thus despite all the shortcomings of the behavioral approach as a general theory of concern or desire, at least on occasion there seems to be a tight connection between concern and dispositions for action, a tighter connection than the cognitive approach seems to be able to countenance, suggesting that the cognitive approach might be going too far in the opposite direction, seeming too representational, too cold. The cognitive account cannot be the account I favor, for obvious reasons, but I believe it should be apparent that it is plagued by problems that can be appreciated quite independently of any particular meta-normative commitments.

Both the conative and the cognitive account experience severe difficulties in handling appraisal, which has led several authors to question whether these are our only options. Indeed, if neither the purely conative nor the purely cognitive account offer a

satisfactory way of capturing appraisal, perhaps the culprit is the underlying division between the cognitive and conative in the first place, which suggests it must be but one or the other. This is precisely the diagnosis of Bennett Helm's (2001), blaming the traditional cognitive-conative divide to undercut all chances of a satisfactory treatment of appraisal, or import, to be forthcoming from the very start. His analysis of the problem of appraisal or import has been extraordinarily helpful in my own attempts to come to terms with the phenomenon, and I mostly share Helm's pessimistic assessment that conative and cognitive accounts alone have the resources to explain import in its full extent. I am less sure, however, whether we ought to follow Helm's rejection of the conative/cognitive divide, as the divide appears virtually logically inevitable, as Michael Smith has convincingly argued (1987), a piece of orthodoxy Helm is keenly aware of. More to the point, however, I suspect Helm's own positive account does not really transcend the distinction itself, but instead only offers a more sophisticated account of the broadly conative sort, which possibly combines some cognitive elements, resulting in complex mental states often called *besires* (J. Altham's term from 1986). Whether this ultimately fully resolves the predicament is unclear, which quickly resurfaces as we turn our attention inside of the relevant complex states, asking how its various pieces are related to each other.

Anyway, and generalizing from the case of emotions that forms his prime target, Helm offers a resourceful solution to how concerns could be at the same time responsive to and constitutive of import or appraisal. Individual concerns are supposed to be embedded within larger patterns of concerns displaying interconnected rational commitments. Helm's central claim is "that import is constituted by a pattern of

evaluative attitudes with the right sort of principle of positive coherence.” (2001:71) His idea of transitional commitments provides a neat example of the kind of rational interconnectedness he is after. The *forward-looking* emotion of hope, for instance that one will secure some appointment, is rationally related to the *backward-looking* emotion of relief, as, to continue the example, when one has learned that one has indeed secured the appointment. Responding to the realization of one’s hopes with relief and contentment seems fitting, whereas responding to it with grief and discontentment would seem unfitting. “The rationality of this pattern means that a failure to feel an emotion one is rationally committed to feeling in virtue of the overall pattern is a failure to respond as one ought to the focus of the pattern.” (2001:72) Because of this, Helm holds “import is *perspectivally subjective* insofar as something is intelligible as having import only from a particular point of view, which is, moreover, relative to the individual. Such a point of view and its relativity to the subject are intelligible in terms of these patterns of emotions. ...Nonetheless ... import is *objective* in the sense that it is a standard in terms of which the warrant of particular emotions is to be assessed and so is conceptual prior to particular emotions.” (2001:73-74) As I understand it, the essential idea of Helm’s proposal is that what we get with overall patterns of emotions and concerns are standards against which particular emotions and concerns can be assessed as inappropriate.

5.9. Appraisal – The Outline of a Solution (*or something close enough*)

In Helm's proposal we can recognize the familiar subjectivist strategy, which accordingly shall be my strategy as well. The strategy takes its point of departure in the recognition that the simple conative account does not work, and in particular that single conative attitudes won't give us what we want, in response to which subjectivists propose raising the number of attitudes, in the hope that many will achieve in concert what one alone was unable to achieve in unison, so long as we manage to interconnect them in the right way. To think it possible that some attitude displays the pertinent receptivity towards certain features of its object is to think it possible that the attitude be appropriate or inappropriate in light of these features. Now we know appropriateness simpliciter is not a condition the subjectivist approach can countenance, but the account may still explain how we can judge the appropriateness of one attitude relative to others, in light of some background norms mediating the interconnection. That a bunch of attitudes alone won't resolve the problem, and consequently that we really need to throw some background norms into the mix is readily apparent once we realize that what we wish to explain is not mere conflict between attitudes but potential *inappropriateness* of some in light of others. If I care for tonight's symphony performance as well as the simultaneous boxing match, I have two concerns that cannot be jointly satisfied, but this does not make one inappropriate in light of the other. In contrast, if I care for the boxing match while also finding myself in the grip of a norm ruling out the enjoyment of violence when more refined sorts of entertainment are available, we now have a norm-concern pair which is at

least formally capable of conferring the status of inappropriateness onto my original pro-attitude towards watching boxing matches. Desires pulling in opposite directions are simply opposing desires, and something else is required to turn mere opposition into normative inappropriateness.

That, in any case, is the outline of the subjectivist strategy, and before I detail it further, let me provide a frank assessment of what I take it to achieve and especially what I don't take it to achieve. Norm-concern complexes or patterns generate the possibility that *particular* attitudes or concerns embedded within larger structures can come to be judged inappropriate on particular occasions, relative to the overall structure, giving rise to an "internal" sense of appropriateness going a considerable way towards capturing the idea of appraisal, yet quite possibly not all the way. What lies beyond that reach I must declare illusory, and hence my above stated pessimism that we won't be able to fully capture import. The remaining gap can be nicely brought out by momentarily returning to McDowell's allusion to secondary qualities as a model for thinking about value (McDowell 1988). McDowell's idea is that secondary qualities display an intrinsic relativity to perceptual states yet that at the same time particular perceptive states could be warranted by the presence of particular secondary qualities, and that the involving circularity is non-vicious. To use the standard example, colors can be both perception-relative and perception-warranting according to McDowell. Now there is one respect in which this idea works and another in which it does not, exemplifying the more general point of how the subjectivist strategy can capture appraisal /import in one respect but not another.

There are two different ways in which a response could be said to be warranted by the presence of some feature, and in particular there are two different correctness/incorrectness contrasts, corresponding to what is often called the internal/external divide, going all the way back to Carnap's discussion on internal and external questions in *Empiricism, Semantics, and Ontology* (in Carnap 1947). First, within our subjective scheme of experiencing color, there is such a thing as experiencing particular colors correctly, where something really is red or blue, and where we may come to have mistaken experience about it. My particular red-sensation could be correct in the sense that it, and not a blue-experience, was called for given my overall color scheme (which others fortunately share). And thus McDowell is correct when he writes (In Sayre-McCord 1988:168) "an object's being such as to look red is independent of its actually looking red to anyone on any particular occasion." But there's another sense in which my particular red-sensation could be said to be called for, namely that there was something in the object meriting the distinctive *colorly* response in the first place, where the focus now is on the contrast of colorly versus non-colorly response, or alternatively, the focus is on the contrast between the traditional colorly response versus some inverted colorly response, a feature, namely REDNESS, which the distinctive colorly response "correctly" responds to and the non-colorly or inverted colorly response miss. And this second sense, I conjecture, is not captured by McDowell's ingenious account, precisely in virtue of his insistence on the specific sensation-relativity of colors. The naïve suspicion about colors is that we might not have evolved colorly-responses to surface-features in the first place, in which case there simply wouldn't have been colors, and that there's really nothing out there we would have missed in that alternative scenario, some feature

calling for or warranting our present colorly scheme as a whole as opposed to some alternative non-colorly scheme or some alternative colorly scheme. Nothing McDowell says can ultimately remove the suspicion that the world as such is not colorful, that color is nothing but a projection, and all McDowell is capable showing is that there are internal rules to the practice of color-perception. Once we are in the color-perceiving business, we encounter internal, quasi built-in correctness-conditions giving rise to the possibility of mistaken particular color sensations. But there is nothing that would explain how being in that business as opposed to being in some different or inverted business is “correct” or “warranted” by how things are, nothing warranting being a color-perceiver in the first place.

Returning to concerns and appraisal, something analogous is the case, as we again encounter two different ways in which concerns could be said to be called for given some situation, and in particular where we encounter two different correctness/incorrectness contrasts vis-à-vis appraisal. First, within our particular concern-pattern, there is such a thing as appraising particular objects correctly. My particular appraisal that something is concern-inviting could be correct in the sense that this particular concern, rather than some contrary concern, was called for given the overall concern pattern encompassing the particular concern. But there’s another sense in which my particular concern could be said to be called for, and the relevant contrast again is best brought out by considering an alternative and inverted concern-pattern, incorporating an altogether different concern-perspective by systematically altering the original’s pro-and con-attitudes. Thus, given some situation, whereas within the first pattern a concerned response is called for, within the second pattern a concern-inverted response is called for. That the subjectivist

approach cannot accommodate appraisal in this more demanding sense, where the original concern correctly appraises while the inverted concern incorrectly appraises what is truly concern-inviting should be clear, prompting a rather straightforward diagnosis of what happens in the subjectivist move. What happens is that by shifting the focus from individual concerns to concern-patterns the subjectivist also shifts the problem from individual concerns to concern-patterns – from how individual concerns may be called for to how entire concern-patterns may be called for, at which stage the subjectivist better proposes ways of coping with the absence of a complete solution than to pretend, as in some old conjuring trick, that his reorientation in focus could suddenly make the problem disappear.

With this important disclaimer, return to the subjectivist strategy for handling appraisal, which is built upon a suitable provision of norms governing the interaction of particular concerns embedded within larger concern patterns. This is a good moment to highlight the remaining element of concern-resilience left out in my previous discussion, namely the element of specificity. The goal is to capture the specific appraisal of some given concern, and for this purpose not every sort of norm grounding conditions for the appropriateness or inappropriateness of the concern will do. The relevant norms must be tailored to the specific way in which the relevant concern appraises its object. The reason why this is such a hard nut to crack is that there are so many different standards of inappropriateness applying to any given concern, many of which appear entirely beside the point with regard to the goal of capturing the specific perspective of appraisal, standards that are not necessarily illegitimate, but too external to offer any help with the present task.

An example will illustrate the problem best, namely the attitude of amusement containing an appraisal of funniness, an attitude at first appearing ill-fit for the format of concern until we think of amusement as a particular way of not being indifferent at funniness, as a concern to be amused by funny things. In any case, there are many norms governing what to be amused by and what not to be amused by that offer no help whatsoever for capturing the specific appraisal of funniness, some of these norms recognizably moral in tone, such as the norm not to find highly offensive jokes amusing. Yet as anyone knows, there are plenty of funny jokes which are morally outrageous, and no civilized person would dare to respond to them with amusement in public. All the same, some of these evil jokes are quite funny, and amusement is the internally fitting response, displaying exactly the right sort of appreciation of funniness characterizing the specific sort of appraisal inherent in amusement. The moralist exhibiting not the slightest hint of amusement at the evil joke is probably a very good person indeed, but also quite probably not of the particularly humorous sort, suffering from some sort of funniness-blindness as it were. Thus, amusement can be justifiably judged inappropriate and off-target in ways that have nothing to do with the specific appraisal we are trying to capture. And with that example I believe the general point becomes apparent at an instance. What we need are not just norms explaining how some concern could be inappropriate, but how it could be inappropriate in the right sort of way.

5.10. The Concern for Recognition

Let me try to put all this together, by addressing one particular instance: The concern for recognition, meaning by *recognition* the notion which is part of the conceptual ballpark containing notions such as approval, praise, esteem, appreciation, and so on, rather than part of the different conceptual ballpark containing notions such as knowledge, belief, representation, awareness, and so on. Recognition resembles a form of honest yet muted praise. The difference between a representational and esteeming sense of recognition, as we may call it, can often be paraphrased in terms of recognition-*to-be* versus recognition-*as*: I recognize Josef Mengele – *Doctor Death* – to be a ruthless medical practitioner, but I do not recognize Mengele *as* a ruthless medical practitioner, since this represents no distinction in virtue of which I could come to hold him in any esteem. It is interesting to note that this ambiguity appears to be a problem of English; the German equivalent for esteeming recognition, *Annerkennung*, displays no such ambiguity, clearly separated from the term *Erkenntnis* denoting representational recognition.

Before outlining standards underwriting the concern for (esteeming) recognition, I need to mention an important disclaimer. What follows is not intended to uncover the nature of recognition as it necessarily is of concern to every rational creature, but pursues instead the more modest goal of lying bare a particularly interesting kind of it, perhaps even a prototypical kind, that is of great concern to (most of) us. This point is even more important to stress since recognition is complex, and no single instance of recognition is an instance of a single kind of recognition; recognition arises on different levels and in

different shapes: for instance, recognition for a deed and recognition for a person are, though conceptually distinct, intimately tied up with each other. My hope in what follows is that the reader will recognize, in a representational manner, a sufficient number of the kinds of observations I unearthed from my own autobiography in his as well, or at least find them helpful in directing his attention to some similar patterns. What matters to my argument is that individual by individual, some similar kinds of norm-concern patterns are discernible in order to support my claim that such patterns can accommodate appraisal to the desired degree, and this end does not require that the same pattern is discernible in each individual, though I rather doubt that we are altogether different in our regard for recognition together with its underwriting principles.

A concern for recognition is one of the most fundamental concerns many of us have, inextricably bound up with our social nature. It's centrality to our identity and dignity has been noted by many authors, many inspired by the writing of Hegel (1988:127-156). I am here simply going to take for granted the profound importance recognition has for (most of) us. The question I wish to focus on is what larger (yet not too large) package of norms and concerns we can identify to ground the right kind of appropriateness conditions for particular recognition-related concern-occasions, patterns explaining how concrete concerns to be recognized on particular occasions may contain an appraisal of the relevant situation as recognition-inviting, patterns explaining, in other words, how concrete concerns may exhibit the sense of responsiveness subjectivists aspire to accommodate too. Part of the answer can be extracted from attending the recognition-complex more closely. Recognition is necessarily discriminatory; not everything can mandate recognition just as not everyone can win. Moreover, when we

care for recognition on particular occasions, it is from *concrete persons* for *concrete accomplishments* we wish to be recognized, and consequently, there is a *who* and a *what*, or a *source* and an *object*, with regard to which specific appropriateness conditions can be identified. Turn first to the question *who* we can sensibly wish to be recognized by.

“Significant others,” Charles Taylor suggests (1992:32), and clearly seems correct in insisting that not just everyone can feature as a proper supplier of recognition, as *a source* of recognition as I chose to call it (add another terminological stipulation: call the recipient of recognition, the person who wishes to be recognized, the *recipient*). In this regard recognition differs from respect, which is owed to everyone, irrespective of familiarity, virtue and accomplishment, and which, unlike recognition, is characteristically non-discriminatory in this regard (respect may be discriminatory in other regards). To be more specific, there are various source-related restrictions, three of which I shall discuss here, beginning with a first class which stems from the rich interaction between our concern for recognition and our concern for the enjoyment of meaningful relationships, a subject broached earlier. People we are completely indifferent towards, with whom we neither enjoy nor seek reciprocal relationships, usually form improper sources of recognition. The explanation why is complex. There seems to be a reciprocal element in recognition, a give and take, and without bestowing concern onto someone expectations for returns in terms of recognition are out of balance. There is also the famous Master-Slave dialectic: the worth the recipient awards to recognition appears bound up with the worth he awards to the source, and indifference towards the latter infects the former with indifference as well – more on this momentarily. In any case, should one nonetheless be accosted by a forceful concern for being recognized by a total

stranger, for instance getting worked up by whether ones distant neighbors recognize one as the great stamp collector one is, then, absent some plausible story, the concern appears internally unfitting, perhaps even pathological, lacking the sufficient tie with a corresponding relationship-related concern. One would wonder: *Why does that matter to you, for neither do you care much about them nor do they care much about you and your stamps?* Notice the required presence of some relationship between source and recipient needs not go very deep. The professional athlete cares for recognition by his anonymous fans; the athlete cannot be totally indifferent towards his fans, however, and still expect his concern for recognition to remain intelligible.

This source-related restriction is further highlighted when we view things from the supplier standpoint whose recognition is sought after. I often find myself in situations (e.g. plane-travel conversations) where I wish my interlocutor would not try so hard to gain my recognition for some pet-project of his. What is called for is tolerance, polite interest, refusal to pass judgment, or some such, but the further request for recognition appears to be going too far. Perhaps it must be admitted that every person at least *potentially* provides an intelligible and proper source of recognition; but if so, the reply would be that it is also true that every person is at least *potentially* a participant of a reciprocal relationship. The emphasis on *potential* really matters, since the famous *strain of involvement* (Strawson 1962) which renders pathological the desire to get personally involved with everyone also renders pathological the desire for recognition from everyone.

It is interesting to note, in this context, that the interaction between the two types of concern, pertaining to recognition and relationship, goes in both directions, where we wish to be recognized by those we care about, but also where our insipient concern for fostering a reciprocal relationship with someone is somewhat tied up with the reasonable expectation to be recognized by them. Again the reciprocal element in recognition neatly explains this. When a lover offers an “anonymous” favor for her beloved subject, sending a poem without return address, for instance, it is usually hoped by the sender that the initial aura of anonymity is dispelled at some point, ideally in an accidental manner unprompted by the lover herself, and it is hoped that the beloved *appreciates* the *favor-done-by-the-lover*. Should the addressee find the poem mediocre at best, an embarrassment in style and composition, it would be safe to predict some (possibly diminishing) effects on the infatuation of the sender. Contrary to the stereotype, love, or at least romantic love, is usually anything but selfless, and one important respect in which this is so has to do with the need to be recognized by those one loves.

Another source-related restriction has to do with expertise and the kind of reverence that is usually associated with it. When a naturalist discovers some peculiarities in the mating behavior of a rare species of birds, recognition by his fellow birders is going to matter incomparably more to him than by people unable of telling apart a sparrow from a finch. Seeking recognition by his peers displaying the appropriate expertise represents an intelligible concern; seeking the same recognition by some know-nothings birdwise does not. A yet further source-related restriction, a hybrid between recognition-related norms pertaining to the *who* and norms pertaining to the *what*, the *object*, has to do with the source’s attitude toward the object of concern. The concern to

be recognized by someone is intelligible only to the extent to which the source does also care for, or could intelligibly be brought to care for, *what* one wishes to be recognized for. Suppose a family-member of an ardent videogame player despises violent computer games. Desiring of his relative to recognize his acquired skill to kill a great number of virtual aliens within short periods of time would seem unfitting; not just would it be unwise and offensive for the player to brag about that skill in the presence of his relative (bragging is an action, an altogether different affair), but it would be unreasonable for the player to *want* his relative to recognize the skill; a verdict that would have been quite different had the concern for recognition been directed towards one of the player's own video-gaming-buddies.

Turn then to the object-side of recognition, the *what* for which we care to be recognized. Recognition-based norms pertaining to the object are no harder to find than those pertaining to the source. First, and rather boringly, the object must exist, or better, the particular states of affairs one desires recognition for must be a fact. One cannot intelligible care to be recognized for what one is not or for what one has not done. A second and more interesting object-related restriction has to do with the recipients' own attitudes towards the object of recognition: the object must matter to them, must be something they care about and are not indifferent towards. Consider a recipient who, regarding some matter as entirely trivial and unnoteworthy, such as his capacity acquired by watching too much TV to detect patterns in which commercials follow which, a capacity he himself considers indicative of bad habits, nothing to be proud of, nonetheless finds in himself a firm concern to be recognized for *that capacity*. This would be completely out of place. Not minding about something and yet minding to be

recognized for it is, if not straightforwardly contradictory, quite out of order. As a consequence, changes in our regard for certain objects must properly be matched by changes in our recognition-related desires as well. It makes no sense for a sprinter to uphold his desire for recognition of his victory after he has learned that his opponents did not try very hard, conspiring perhaps to let him win the race to lift up his spirits after a series of losses; or for a doctor for curing a famous patient after she has learned that by doing so she only gave the dictator more days of practicing cruelly; or for the researcher for synthesizing a virus after she has learned that thereby she caused a pandemic.

The interaction between ones regard for some object and ones desire to be recognized for it is usually subtle and evolves gradually over time. Consider a filmmaker who desires to be recognized as an artist, and not merely as an entertainer. Then, driven by that desire into the arts scene, and gaining greater exposition to it, he comes to find it difficult to place his work in any conversation with what he finds there. Moreover, coming to regard their preoccupations as overblown and out of touch, his initial grand esteem for the higher arts slowly fades. Finally he (mostly) contents himself with being an entertainer and to be recognized only as such. And yet the development of his desire might not yet have caught up, where there remains a small residue of a desire for recognition as an artist which he cannot fully expunge, which, given his new views about the arts are really sincere and wholehearted, he himself must regard as a glitch, a regrettable failure to escape the shadow of a discredited evaluative framework.

Somewhat harder to document, though quite plausible, is a third and cultural element guiding our concern for recognition, a natural extension of the fact that both

source and recipient must hold a positive outlook on the object of recognition. Usually, though not always, appreciation of the object of recognition must be more widely shared, by one's peers, by one's social group, or even by one's society at large. This could be so in part because the source comprises a multitude of people; but often the social dimension seems to go beyond that, providing a background against which particular concerns for recognition become intelligible. Consider recognition for being a great scholar. Being a scholar designates a recognition-apt social type in terms of which even people with little sense for scholarship can find the desire for recognition intelligible. Their applause in a ceremony awarding an academic distinction to a scholar for her excellent work on late Roman poetry makes sense even though late Roman poetry means absolutely nothing to them. It is part of a larger cultural context distinguishing some but not other activities and pursuits as worthy of recognition.

The example of recognition can also serve to highlight norm-irrelevancy in addition to norm-relevancy with regard to capturing appraisal, documenting the force of the cautionary notes from above pertaining to the diversity and particularity of concerns. Suppose a stoic, striving to gain maximal independence from the opinion of others, seeks to expunge or disown all desire for recognition. In the struggle between a desire for social independence and for social dependence and involvement, seeing no easy reconciliation, the stoic resolves to firmly come down on the side of independence at the total expense of social involvement. Still, his general contempt of recognition need not diagnose a recognition-specific fault in his particular desires for recognition on particular occasions, which I doubt even a stoic can fully silence. The stoic, too, should admit that there are proper and improper cases of recognition *qua recognition*, cases that correctly and

incorrectly appraise their object as fit for recognition, but only decries recognition as a discredited phenomenon (nice try!). Emperor Nero was internally mistaken to sing his own praise while burning down much of Rome, a deed calling forth no praise, while Mark Aurelius' victories in Germania were fit for recognition. Or suppose, to take a second example, a Christian despising pride as the most serious of the seven deadly sins, considering, moreover, the desire for recognition to be nothing but a dressed up form of pride. Still, such general condemnation of recognition does not necessarily invalidate any particular case of recognition *as* a proper case of recognition, but only invalidates recognition for external and quite recognition-unrelated reasons. The Christian too should admit that though sinful, Michelangelo's desire for recognition for his extraordinary decorations of the Sistine Chapel is internally fitting; in stark contrast to had Michelangelo sought recognition for his abrasive character. In this latter case Michelangelo would have been guilty on two counts, not just one: harboring a desire both sinful and unfitting. Thus, while the stoic and the Christian consider their reservations about recognition legitimate, they must also acknowledge that there is a sense in which those reservations do not really speak to recognition as such.

I have thus outlined three source-related and three object-related norms governing the concern for recognition, none of which provide necessary conditions, all of which are negotiable with regard to their details, and yet when they apply in concert, they seem to give us something of a paradigmatic case of a proper concern for recognition. Suppose I have achieved something truly astonishing which profoundly matters to me, my peers, and my friends. I am fully invested in the relevant project and people. With all the right sort of connections in place between my concern for recognition, my other concerns, as

well as other people and their concerns, the concern presents itself as responsive to an invitation for recognition, containing an appraisal of its object as fit for recognition. This hopefully adds plausibility to my claim that the complex norm-concern cluster underwriting recognition has the capacity of accommodating the appraisal internal to my concrete concern of recognition, providing an instance for the more general claim that complex norm-concern clusters can accommodate appraisal.

This concludes my discussion of concerns, and my discussion of practical deliberation consistent with meta-normative skepticism. I presented a revisionary account of deliberation as a very sophisticated form of norm-guided motivation with the aim of accommodating core features of practical deliberation. I first rejected the fictionalist proposal for how to respond to skepticism as inferior to the revisionist alternative, arguing that fictionalism, once fully worked out, is going to lead us right to the very doorsteps of revisionism, in the process rendering superfluous the twisted fictionalist attitude of make-believing in a reality truly disbelieved. I then presented the challenge that without a normative account there cannot be an adequate deliberative account, the challenge, in particular, that revisionary subjectivism cannot accommodate the notion of correct deliberation essential to any deliberation worth the name. To this I responded by aiming to show how the subjectivist can accommodate correctness-permitting deliberation from within, by helping himself to norms underwriting the deliberative process, norms void of normative authority for sure yet nonetheless generating formal correctness conditions which, in conjunction with motivational force acquired through our commitments, enable a form of practical deliberation incorporating internal correctness and incorrectness conditions. Thereby I presented deliberation as a very

sophisticated form of motivational capacity, a move prone to raise some eyebrows, calling for the section trying to justify the subjectivist's turn to the motivational.

The second half of the chapter was dedicated to a close discussion of concerns, one core component engendering deliberation. Attending to concerns more closely revealed an astonishing degree of complexity and richness, which, despite the long noticed importance of desires and concerns for practical deliberation, has not always received adequate treatment in the literature. Concerns are unruly fellows, displaying an intriguing set of properties including diversity, specificity, particularity, and synchronic and diachronic stability. Moreover, concerns appeared to contain an element of appraisal, where there often seems, in the relevant object, an element of invitation *drawing in* the concern, such that concern do not present themselves as a mere blind reaction aroused by certain aspects of their objects, but as a receptivity to their *attractiveness*, which in turn explains why it might be possible to construe the concern as something occasionally *called-for by the attractiveness of the relevant feature*, as something that can be *appropriate or inappropriate* depending on whether the feature is in fact attractive. In light of this, my goal was to accommodate the element of appraisal in a manner consistent with skepticism. The final section then presented a case study of concern – the concern for recognition – designed to illustrate how the subjectivist can capture appraisal.

References

- Anscombe, G. E. M. [1967], *Intention*, Ithaca, NY: Cornell University Press.
- Arpaly, Nomy [2006], *Merit, Meaning, and Human Bondage*, Princeton: Princeton University Press.
- Bedke, Matthew [2010], *Might All Normativity Be Queer?*, Australasian Journal of Philosophy 88(1): 41-58.
- Berlin, Isaiah [2000], *The Proper Study of Mankind, An Anthology of Essays*, New York: Farrar, Straus and Giroux.
- Blackburn, Simon [1998], *Ruling Passions*, Oxford: Oxford University Press.
- Brandom, Robert B. [1994], *Making It Explicit*, Cambridge, MA.: Harvard University Press.
- Brandon, E. P. [1980], *Subjectivism and Seriousness*, The Philosophical Quarterly, Vol. 30, No. 119:97-107.
- Brandt, Richard [1979], *A Theory of the Good and the Right*, Oxford: Oxford University Press.
- Bratman, Michael E. [1999], *Faces of Intention*, Cambridge, UK; Cambridge University Press.
- Bratman, Michael E. [2000], *Reflection, Planning, and Temporally Extended Agency*, The Philosophical Review, Vol. 109, No. 1:35-61.

Bratman, Michael [2003], *A Desire of One's Own*, Journal of Philosophy, Volume C, No.5, May, 221-242.

Brink, David [1989], *Moral Realism and the Foundation of Ethics*, Cambridge, UK: Cambridge University Press.

Broome, John [1999], *Normative Requirements*, Ratio 12: 398–419.

Calhoun, Cheshire [2009], *What Good Is Commitment?*, Ethics 119:613-641.

Carnap, Rudolf [1947], *Meaning and Necessity*, Second Edition, Chicago: University of Chicago Press.

Chalmers, David J. [2009], *Ontological Anti-Realism*, in: Chalmers, Manley & Wasserman 2009.

Chalmers, David & Manley, David & Wasserman, Ryan (eds.) [2009], *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press.

Copp, David [1995], *Morality, Normativity, and Society*, Oxford: Oxford University Press.

Copp, David [2007], *Morality in a Natural World*, Cambridge: Cambridge University Press.

Cullity, Garrett & Gaut, Berys [1997], *Ethics and Practical Reason*, Oxford: Oxford University Press.

Cuneo, Terence [2007] *The Normative Web*, Oxford: Oxford University Press.

Dancy, Jonathan [2000], *Practical Reality*, Oxford: Oxford University Press.

Dancy, Jonathan [2004a], *Ethics Without Principles*, Oxford: Oxford University Press.

Dancy, Jonathan [2004b], *Discussion – On Knowing What One is Doing*, *Philosophical Studies* 121: 239–247.

D’Arms, Justin & Jacobson, Daniel [2000], *Sentiment and Value*, *Ethics* 110:722-748.

Darwall, Stephen L. [1992], *Internalism and Agency*, *Philosophical Perspectives*, Vol. 6, *Ethics*:155-174.

Dennett, Daniel [1987], *The Intentional Stance*, Cambridge, MA: MIT Press.

Dreier, James [2001], *Humean Doubts about Categorical Imperatives*, in: *Millgram* 2001:27-47.

Dworkin, Ronald [1996], *Objectivity and Truth: You'd Better Believe it*, *Philosophy and Public Affairs*, Vol. 25, No. 2., pp. 87-139.

Elster, Jon [1985], *Ulysses and the Sirens*, Cambridge, UK: Cambridge University Press.

Enoch, David [2006], *Agency, Shmagency: Why Normativity Won't Come from What is Constitutive of Agency*, *Philosophical Review* 115, 169-198.

Enoch, David [2007], *An Outline of an Argument for Robust Metanormative Realism*, *Oxford Studies in Metaethics* 2 (2007), pp. 21-50.

Enoch, David [forthcoming a], *Taking Morality Seriously: A Defense of Robust Realism*, forthcoming Oxford University Press.

Enoch, David [forthcoming b], *Reason-Giving and the Law*, forthcoming in *Oxford Studies in the Philosophy of Law*.

Enoch, David [forthcoming c], *Not Just a Truthometer: Taking Oneself Seriously (but not too seriously) in Cases of Peer Disagreement*, forthcoming in *Mind*.

Ferrero, Luca, *Constitutivism and the Inescapability of Agency*, in Russ Shafer-Landau 2009:303-333.

FitzPatrick, William [2005], *The Practical Turn In Ethical Theory: Korsgaard's Constructivism, Realism, and the Nature of Normativity*, *Ethics* Vol. 115, No. 4 (July, 2005).

Frankfurt, Harry [1982], *The Importance of What We Care About*, *Synthese*, Vol. 53, No. 2, *Matters of the Mind* (November), pp. 257-272.

Frankfurt, Harry [1988], *The Importance of What We Care About*, Cambridge, UK: Cambridge University Press.

Foot, Philippa [1972], *Morality as a System of Hypothetical Imperatives*, *Philosophical Review* Vol. 81, No. 3. (Jul., 1972), pp. 305-316.

Gell-Mann, Murray [2007], TED Lecture on Beauty and Physics:
http://www.ted.com/index.php/talks/murray_gell_mann_on_beauty_and_truth_in_physics.html

Gert, Joshua [2004], *Brute Rationality*, Cambridge, UK: Cambridge University Press.

Gibbard, Allan [1990], *Wise Choices, Apt Feelings*, Cambridge, MA: Harvard University Press.

Gibbard, Allan [2003], *Thinking How to Live*, Cambridge, MA: Harvard University Press.

Gilbert, Daniel [2007], *Stumbling on Happiness*, New York: Vintage Books.

Glüer, K & Wikforss [2009], *Against Content Normativity*, *Mind* 2009 118(469):31-70.

Griffin, James [1986], *Well-Being, Its Meaning, Measurement, and Moral Importance*, Oxford: Oxford University Press.

Griffin, James [1996], *Value Judgement*, New York: Oxford University Press.

Guyer, Paul [2006], *Kant*, New York: Routledge.

Haldane, J.B.S. [1927], *Possible Worlds and Other Essays*, London: Chatto and Windus (Reprint 1932).

Hampton, Jean E. [1998], *The Authority of Reason*, Cambridge, UK: Cambridge University Press.

Hart, H.L.A. [1955], *Are there any Natural Rights?*, *Philosophical Review* Vol. 64, No. 2 (Apr., 1955), pp. 175-191.

Hart, H.L.A. [1994], *The Concept of Law*, Oxford: Clarendon Press.

Hegel, G.W.F. [1988], *Phänomenologie des Geistes*, Hamburg: Meiner Verlag.

Hieronymi, Pamela [2005], *The Wrong Kind of Reason*, Journal of Philosophy, Vol. cii, no. 9, September 2005.

Hume, David [2000], *A Treatise of Human Nature*, New York: Oxford University Press.

Jackson, Frank [1998], *From Metaphysics to Ethics*, New York: Oxford University Press.

Jackson, Pettit, Smith [2004], *Mind, Morality and Explanation*, Oxford: Oxford University Press.

Joyce, Richard [2001], *The Myth of Morality*, Cambridge, UK: Cambridge University Press.

Joyce, Richard [2005], *Moral Fictionalism*, in: Kalderon 2005:287-313.

Joyce, Richard [2006], *The Evolution of Morality*, Cambridge, MA: MIT Press.

Kagan, Shelly [1988], *The Additive Fallacy*, Ethics 99:5-31.

Kalderon, Mark Eli [2005], *Fictionalism in Metaphysics*, Oxford: Oxford University Press.

Kant, Immanuel [1998], *Grundlegung zur Metaphysic der Sitten*, Darmstadt, Germany: Wissenschaftliche Buchgesellschaft.

Kim, Jaegwon [1998], *Mind in a Physical World*, Cambridge, MA, MIT Press.

Kolodny, Nico [2005], *Why Be Rational?*, Mind 114:455: 509–63.

Korsgaard, Christine [1996], *The Sources of Normativity*, Cambridge, UK: Cambridge University Press.

Korsgaard, Christine [1996b], *Creating the Kingdom of Ends*, Cambridge, UK: Cambridge University Press.

Korsgaard, Christine [1996c], *Two distinctions in goodness*, in: Korsgaard 1996b.

Korsgaard, Christine [1997], *The Normativity of Instrumental Reason*, in: Cullity & Gaut 1997:215-254.

Korsgaard, Christine [2008], *The Constitution of Agency, Essays on Practical Reason and Moral Psychology*, Oxford: Oxford University Press.

Korsgaard, Christine [2009], *Self-Constitution, Agency, Identity and Integrity*, New York: Oxford University Press.

Lewis, David [1989], *Dispositional Theories of Value*, Proceedings of the Aristotelian Society, Supplementary Volumes, Vol. 63, pp.113-137.

Lewis, David [2002], *Convention*, Oxford: Blackwell.

Loeb, Don [2003], *Gastronomic Realism – A Cautionary Tale*, The Journal of Theoretical and Philosophical Psychology.

Mackie, J. L. [1977], *Ethics – Inventing Right and Wrong*, New York: Pelican 1977.

Marmor, Andrei [2008], *The Nature of Law*, The Stanford Encyclopedia of Philosophy, Edition Fall 2008.

Marmor, Andrei [2009], *Social Conventions*, Princeton: Princeton University Press.

McDowell, John [1988], *Value and Secondary Qualities*, in: Sayre-McCord 1988:166-180.

Millgram, Elijah [2001], *Varieties of Practical Reasoning*, Cambridge, MA: MIT Press.

Moran, Richard [1988], *Making Up Your Mind: Self-Interpretation and Self-Constitution*, Ratio No. 1 (1988): pp.135–51.

Nagel, Thomas [1970], *The Possibility of Altruism*, Princeton University Press.

Nagel, Thomas [1979], *Mortal Questions*, Cambridge, UK: Cambridge University Press.

Nagel, Thomas [1986], *The View from Nowhere*, New York: Oxford University Press.

Nichols, Shaun [2004], *Sentimental Rules*, New York: Oxford University Press.

Nolan, Daniel [2005], Fictionalist Attitudes about Fictional Matters, in: Kalderon 2005:204-233.

Nussbaum, Martha C. [2001], *Upheavals of Thought*, Cambridge, UK: Cambridge University Press.

Parfit, Derek [1984], *Reasons and Persons*, Oxford: Oxford University Press.

Parfit, Derek [2006], *Normativity*, in: Shafer-Landau 2006.

Parfit, Derek [2011], *On What Matters*, Oxford: Oxford University Press.

Pettit, Philip, and Smith, Michael [1990], *Backgrounding Desires*, *Philosophical Review*, 99, 565-92.

Prinz, Jesse [2007], *The Emotional Construction of Morality*, Oxford: Oxford University Press.

Quine, W.V. & Ullian, J.S. [1978], *The Web of Belief*, New York: McGraw-Hill.

Railton, Peter [2003], *Facts, Values and Norms*, Cambridge, UK: Cambridge University Press.

Rawls, J. [1955], *Two Concepts of Rules*, *Philosophical Review* 64: 3-32.

Rawls, John [1971], *A Theory of Justice*, Cambridge, MA: Harvard University Press.

Rawls, John [2005], *Political Liberalism*, New York: Columbia University Press.

Raz, Joseph [1999], *Practical Reasons and Norms*, Oxford: Oxford University Press.

Raz, Joseph [2010], *Reason, Reasons & Normativity*, in: Shafer-Landau 2010.

Richardson, Henry [1994], *Practical Reasoning about Final Ends*, Cambridge, UK: Cambridge University Press.

Rosati, Connie [2003], *Agency and the Open Question Argument*, *Ethics* 113: 490-527.

Rosen, Gideon [1994], *Objectivity and Modern Idealism: What is the Question?*, in: Michael & Leary-Hawthorne (ed.), *Philosophy of Mind, The Place of Philosophy in the Study of Mind*, Dordrecht: Kluwer Academic Publishers.

Sayre-McCord, Geoffrey, ed. [1988], *Moral Realism*, Ithaca, NY: Cornell University Press.

Sayre-McCord, Geoffrey [1997], The Metaethical Problem, *Ethics* 108:55-83.

Scanlon, Thomas [1998], *What We Owe to Each Other*, Cambridge, MA: Harvard University Press.

Scanlon, Thomas [2008], *Moral Dimensions*, Cambridge, MA: Harvard University Press.

Scheffler, Samuel [1992], *Human Morality*, Oxford: Oxford University Press.

Schneewind, J. B. [1998], *The Invention of Autonomy*, Cambridge, UK: Cambridge University Press.

Schroeder, Mark [2008a], *The Slave of Passions*, Oxford: Oxford University Press.

Schroeder, Mark [2008b], *Having Reasons*, *Philosophical Studies* 139(1): 57-71, April 2008.

Searle, J. R. [1969], *Speech Acts*, Cambridge, MA: Cambridge University Press.

Searle, John R. [1995], *The Construction of Social Reality*, New York: The Free Press.

Setiya, Kieran [2007], *Reasons without Rationalism*, Princeton, Princeton University Press.

Shafer-Landau, Russ [2003], *Moral Realism – A Defense*, Oxford: Oxford University Press.

Shafer-Landau, Russ (ed.) [2006], *Oxford Studies in Metaethics, Volume I*, Oxford: Oxford University Press.

Shafer-Landau, Russ (ed.) [2007], *Oxford Studies in Metaethics, Volume II*, Oxford: Oxford University Press.

Shafer-Landau, Russ (ed.) [2008], *Oxford Studies in Metaethics, Volume III*, Oxford: Oxford University Press.

Shafer-Landau, Russ (ed.) [2009], *Oxford Studies in Metaethics, Volume IV*, Oxford: Oxford University Press.

Shafer-Landau, Russ (ed.) [2010], *Oxford Studies in Metaethics, Volume V*, Oxford: Oxford University Press.

Shah, Nishi [2003], *How Truth Governs Belief*, *The Philosophical Review*, Vol. 112, No. 4 (October 2003), pp.447-482.

Shah, Nishi & Velleman, J. David [2005], *Doxastic Deliberation*, *The Philosophical Review*, Vol. 114, No. 4 (October 2005), pp.497-534.

Sher, George [1987], *Desert*, Princeton: Princeton University Press.

Smith, Michael [1987], *The Humean Theory of Motivation*, *Mind*, 96:36-61.

Smith, Michael [1994], *The Moral Problem*, Oxford: Blackwell.

Smith, Michael [2004], *Ethics and the A Priori*, Cambridge, UK: Cambridge University Press.

Stern, Robert [2000], *Transcendental Arguments and Skepticism*, New York: Oxford University Press.

Stich, Stephen P. [1990], *The Fragmentation of Reason*, Cambridge, MA: MIT Press.

Strawson, Peter [1962], *Freedom and Resentment*, Proceedings of the British Academy, 48:1-25.

Street, Sharon [2008], *Constructivism about Reasons*, in: Shafer-Landau 2008.

Stroud, Barry [1968], *Transcendental Arguments*, Journal of Philosophy, 65, pp.241-52.

Taylor, Charles [1989], *Sources of the Self, The Making of Modern Identity*, Cambridge, MA: Harvard University Press.

Taylor, Charles [1992], *Multiculturalism and the Politics of Recognition*, Princeton: Princeton University Press.

Taylor, Charles [1977], *What is human Agency?*, in T. Mischel (ed.) *The Self*, Oxford, Blackwell, pp.103-35.

Taylor, Richard, *Good and Evil*, Amherst, New York: Prometheus Books.

Timmons, Mark [1999], *Morality without Foundation*, New York: Oxford University Press.

Thompson, Judith Jarvis [2007], *Normativity*, in: Shafer-Landau 2007.

Velleman, J. David [2000], *The Possibility of Practical Reason*, Oxford: Oxford University Press.

Velleman, J. David [2004], *Replies to Discussion on The Possibility of Practical Reason*, *Philosophical Studies* 121:277-298.

Velleman, J. David [2009], *How We Get Along*, Cambridge, UK: Cambridge University Press.

Wallace, R. Jay [1994], *Responsibility and the Moral Sentiments*, Cambridge, MA: Harvard University Press.

Watson, Gary [1975], *Free Agency*, *Journal of Philosophy* LXXII, 8 (April 24):205-20.

Wedgwood, Ralph [2007], *The Nature of Normativity*, Oxford: Oxford University Press.

Williams, Bernard [1985], *Ethics and the Limits of Philosophy*, Cambridge, MA: Harvard University Press.

Williams, Bernard [1981], *Moral Luck*, Cambridge, UK: Cambridge University Press.

Williams, Bernard [1981b], *Internal and External Reasons*, in: Williams 1981:101-113.

Williams, Bernard [1995], *Making Sense of Humanity*, Cambridge, UK: Cambridge University Press.

Williams, Bernard [1995b], *Internal Reasons and the Obscurity of Blame*, in Williams 1995:35-45.

Wright, Crispin [1992], *Truth and Objectivity*, Cambridge, MA: Harvard University Press.